
Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**

Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 ; Search time 5830.82 seconds
(without alignments)
6486.799 Million cell updates/sec

Title: US-09-763-292-1

Perfect score: 1746

Sequence: 1 cttcaatggatccctttgtg.....acatattattataaataaga 1746

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues

Total number of hits satisfying chosen parameters: 44045606

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096F_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US098_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US099_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6033_MERGED_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US10_MERGED_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US08_MERGED_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US08_MERGED_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/PCT_MERGED_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1745	100.0	1745	30	US-09-763-292-1
2	1739.6	99.6	1745	3	US-07-864-962C-2
3	1739.6	99.6	1745	6	US-08-238-821A-2
4	1739.6	99.6	1745	14	US-09-023-655-1059
5	1739.6	99.6	1745	31	US-09-880-107-2370
6	1739.6	99.6	1745	55	US-60-226-176-749
7	1739.6	99.6	1745	56	US-60-233-468-749
8	1739.6	99.6	1745	64	US-60-313-371-749
9	1739.6	99.6	1745	18	US-09-488-127-58
10	1739.6	99.6	1748	18	US-09-488-127-58
11	1625.2	93.1	1940	27	US-09-698-010-12366
12	1534.6	87.9	1845	14	PCT-US98-16979-4
13	1534.6	87.9	1845	15	US-09-023-655-1057
14	1534.6	87.9	1845	15	US-09-135-296-4
15	1534.6	87.9	1845	55	US-60-226-176-753
16	1534.6	87.9	1845	56	US-60-233-468-753
17	1534.6	87.9	1845	64	US-60-313-371-753
18	1534.6	87.9	1852	3	US-07-864-962C-10
19	1534.6	87.9	1852	6	US-08-238-821A-10
20	1534.6	87.9	1857	18	US-09-471-275-6
21	1529.8	87.6	1854	3	US-07-864-962C-4
22	1529.8	87.6	1854	6	US-08-238-821A-4
23	1529.8	87.6	1854	14	US-09-023-655-1056
24	1529.8	87.6	1854	31	US-09-880-107-2371
25	1512.6	86.6	1916	50	US-60-172-373-8646
26	1466.6	84.0	1473	55	US-60-226-176-750
27	1466.6	84.0	1473	55	US-60-233-468-750
28	1466.6	84.0	1473	64	US-60-313-371-750
29	1458.6	83.5	1473	6	US-08-277-031A-15
30	1458.6	83.5	1473	5	US-08-277-031B-15
31	1453.2	83.2	1835	54	US-60-213-360-5066

Db 721 TTATGAAAGTGATATTTTGGAGAAAGTAAAGAACACCAAGATCGATGGACATCAACA 780
QY 781 acctcggagactttattgatttcttctgatacaaatggagagaaagcaaaccaaac 840
Db 781 ACCCTCGGACTTATTTGATTCCTTCTGATCAAAATGGAGAGAAAGCAAAACCAAC 840
QY 841 agtctgaattcactattgaaacttggtaactcactgcagctacttacttggagctgga 900
Db 841 AGTCTGAATTCATATTGAAACTTGGTAATCAGTGCAGCTGACTTACTTGGAGCTGGGA 900
QY 901 cagagacaagcagcaaccctgagatatgctctctctctctctctctgctggaagcaccagag 960
Db 901 CAGAGACAAGCAACACCTTGAGATATGCTCTCTCTCTCTGCTGAAGCACCCAGAGG 960
QY 961 tcacagtaaatccaggaagagattgaactgctgctgctgctgctgctgctgctgctgca 1020
Db 961 TCACAGCTAAAGTCCAGGAAGATGGAAGTGTCTATTGGCAGAAACCGAGCCCTGCA 1020
QY 1021 tgcacagaggggacacatgcccctacacagatgctgctgctgctgctgctgctgctgca 1080
Db 1021 TGCAGGACAGGGGCCACATGCCCTACACAGATGCTGTGTGTCAGAGGTCCAGATACA 1080
QY 1081 tggactctatccccaccagcctgccccatgcagtgacctgtgacctgtaaatcagaact 1140
Db 1081 TCGACCTCATGCCACAGCCTGCCCATGCAGTGCAGTGCAGCTGACGTTAAATTCAGAACT 1140
QY 1141 acctcattccccagggcaacacatattaaacttccctcacttctgtgtctacatgacaaca 1200
Db 1141 ACCTCAITCCCAGGGCAACACCATATTAACTTCGCTCACITTTCTGTGTACATGACAACA 1200
QY 1201 aagaatttcccaaccagagatgtttgacctcgtcacttctgctgctgctgctgctgctgca 1260
Db 1201 AAGAAATTTCCCAACCAGAGATGTTGACCCCTCGTCACITTTCTGGATGAAGGTGGAAT 1260
QY 1261 ttaagaaagtaactacttctgctgctgctgctgctgctgctgctgctgctgctgctgca 1320
Db 1261 TTAAGAAAGTAACTACTTCTATGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1320
QY 1321 gctgcccgcagtgagctgttttattctgaccttatttctacagactttaaactgca 1380
Db 1321 GCCTGCCCGCATGGAGCTGTTTTATTCTGACCTTCAITTTACAGAACTTTAACTTGA 1380
QY 1381 aatctctgattgacccaaaggacctgacacaaactcctgtgtgctgctgctgctgctgctg 1440
Db 1381 AATCTCTGATTGACCCAAAGGACCTTGACACAACCTCTGTTGCTCAATGGATTGCTTCTG 1440
QY 1441 tccgcctctctacagctgtgcttctcattctgctgctgctgctgctgctgctgctgctg 1500
Db 1441 TCCGCCCCCTTATCAGCTGTGCTTCAITTCCTGCTGAGAGACAGATGGGTCTGGCTG 1500
QY 1501 ctctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1560
Db 1501 CTCTGTGCTGCTCCCTGCGAGCTCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
QY 1561 tcttctgacctgctcctacaaatttcccttcccttcccttcccttcccttcccttcccttcc 1620
Db 1561 TCTTCTGACCCGCTCATCTCACATTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCT 1620
QY 1621 cattaaaaagtttccactgcaaatatctgcttcttcccttcccttcccttcccttcccttcc 1680
Db 1621 CATTAATAAAGTTTCTACGTGCAATATATCTGCTATTTCCCTTACTCTATTAATAGTAC 1680
QY 1681 attgagtcacataatgctgatacttcttaattgagttgagttgagttgagttgagttgagtt 1740
Db 1681 ATTGAGTGCACATATGCTGATACTTGTCTAATGTTGAGTTGAGTTGAGTTGAGTTGAGTT 1740
QY 1741 aataga 1746
Db 1741 AATAGA 1746

RESULT 4

US-09-023-655-1059

; Sequence 1059, Application US/09023655
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1059:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1746 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g181343
; US-09-023-655-1059

Query Match 99.6%; Score 1739.6; DB 14; Length 1746;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctccaatgagctcttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 60
Db 1 CTTCANTGGATCCCTTTTGGTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 60
Qy 61 tctggagacagactctggagagagaaactccctctctggccccactcctcctcctcagtgga 120
Db 61 TCTGGAGACAGAGCTCTGGGAGAGGAAACTCCCTCTCTGGCCCCACTCCTCTCCCACTGA 120
Qy 121 ttgaaatctctacagataatattaagagatgacgaaatccttaaccaatctctcaaa 180
Db 121 TTGGAATATCTACAGATAGATATTAAGGATGTCAGGAATCTCTTAACCAATCTCTCAA 180
Qy 181 aaatatggccctgtgttctactctgtattttggcctggaacgcatggtggtgctgcatg 240
Db 181 AAATCTATGGCCCTGTGTCTACTCTGTATTTTGGCCTGGACGCATGTTGTGTGCTGATG 240
Qy 241 gataatgagtgaggaagccctgattgatcttggagagagagtttcttggaagagcc 300
Db 241 GATATGAAGTGTGAAGGAAGCCCTGATGTGATCTTTGGAGAGAGTCTTTCTGGAGAGGCC 300
Qy 301 atttccccactgagctgaagaactaacagagatttggagatttggagatttgcagcaatgaaaga 360
Db 301 atttccccactgagctgaagaactaacagagatttggagatttggagatttgcagcaatgaaaga 360

```
Db 301 ATTTCACCTGGCTGAAGAGCTAAACAGAGGATTTGGAATCGTTTTTCAGCAATGGAAGA 360
Qy 361 gatgaagaggaattcgggtttccctcatgacgtcggaatttttggatggggaaga 420
Db 361 GATGGAAGGAGATCCGGCGTTTCTCCCTCATGACGCTGGGAATTTTGGATGGGGAAGA 420
Qy 421 ggaacattgaggaccgtgttcaagagagaagaccgctgacctgtggaagggttgagaaaa 480
Db 421 GGACCATTTAGGACCGTGTTCAGAGGAGCCCGCTGCCCTTGTGGAGGAGTTGAGAAAAA 480
Qy 481 ccaaggcttccacctgtgatcccaatttcacctggggtgtgctcctcgcaatgtgatct 540
Db 481 CCAAGGCTTCACCGTGTGATCCACATTTTCATCTGGGCTGTCTCCCTGCAATGTGATCT 540
Qy 541 gctccattatttccagaacggtttcgattataaagatcacaaatttcttaacttgatgg 500
Db 541 GCTCCATTAATTTCCAGAAACGTTTCGATATAAAGATCAGCAATTTCTTAACCTTGATGG 500
Qy 601 aaaaattgaatgaataacatcagattgtgaagcacccccctggatccagatatgcaataatt 560
Db 601 AAAAAATTGAATGAATAACATCAGGATTTGAAGCACCCCTGGATCCAGATATGCAATAATT 560
Qy 661 ttccactatcattgatttatttccgggaaccatacaaaattacttaaaacccttgctt 720
Db 661 TTCCCACTATCATTTGATTGATTTATTTCCCGGGAACCCATAAACAATTAATTAACCTTTGCTT 720
Qy 721 ttatgaaatgatatttttggagaagtaaaagacacaccaaataatcgatggacatcaaca 780
Db 721 TTATGAAAGTGATATTTTGGAGAAAGTAAAGACACCCAGATTCGATGGACATCAACA 780
Qy 781 accctcgggaacttattgatttctctctgatcaaaatgagaggaagaaagcaaaacccaac 840
Db 781 ACCCTCGGACCTTTATTTGATTGCTTCTCTGATCAAAATGAGAGGAAAGCAAAACCAAC 840
Qy 841 agtctgaattcactatgaaaacttggtaactcactcgactgaacttacttgagactgga 900
Db 841 AGTCTGAATTCATTTGAAACTTGGTAATCACTGCAGCTGACTTACTTTGGAGCTGGGA 900
Qy 901 cagagacaagaagcaaaccttgatgatgctctctctctctctctctctctctctctctct 960
Db 901 CAGAGACAAGAGCAACACCTTGAGATATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Qy 961 tcacagtaaatccaggaagagattgaactgtctatgttgagagaacccggagcccttga 1020
Db 961 TCACAGCTAAAGTCCAGGAAGAGATTGAACGGTGTCAITTTGGCAGAAACCGGAGCCCTGCA 1020
Qy 1021 tgcacagagggggccacatgcctcacacagatgctgtgtgtgcacagaggtccagagataca 1080
Db 1021 TGCAGSACAGGGGCCACATGCCCTACACAGATGCTGTGTGTGCACAGGTCCAGAGATACA 1080
Qy 1081 tcgacctatccccaccagctgcccctgacgtgacgtgacgtgacgttaaatccagaact 1140
Db 1081 TCGACTCATCCCAACAGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCT 1140
Qy 1141 acctcattcccaaggagcaaacattattaaacttccctcacttctgtgtgtacatgacaaca 1200
Db 1141 ACCTCATTTCCCAAGGCAACACCATTAATTAATTTCCCTCAGCTTCTGTCTACATGACAAACA 1200
Qy 1201 aagaatttcccaaccagagatttgaacctgtcacttctctgtatggaagggtggaatt 1260
Db 1201 AGAATTTCCCAACCCAGAGATTTTGACCTCGTCACTTTCTGTGATGAAGGTGGAATTT 1260
Qy 1261 ttaagaagaataactacttcatgcttctctcagcagaaaaacagatttgtgtggagagg 1320
Db 1261 TTAAGAAAGTAACTACTTACTTCTCTCAGCAGGAACCGATTTGTGTGGAGAGG 1320
Qy 1321 gcttggcccgatggagctgttttatctcgtaccttattttacagaactttaacctga 1380
Db 1321 GCTTGGCCCGATGGAGCTGTTTTTATTTCTGTGACCTTCAITTTACAGAACTTTTAACCTGA 1380
Qy 1381 aatctctgattgaccaaaaggacctgaacaaactctatttgcattgcaatggattgcttctg 1440
Db 1381 AATCTCTGATTGACCCAAAGGACCTTGACACAACCTCTCTGTGTTGCAATGGATTTGCTTCTG 1440
```

RESULT 5

US-09-880-107-2370

: Sequence 2370, Application US/09880107

: GENERAL INFORMATION:

: APPLICANT: Horne, Darci T.

: APPLICANT: Vockley, Joseph G.

: APPLICANT: Scherif, Uwe

: APPLICANT: Gene Logic, Inc.

: TITLE OF INVENTION: Gene Expression Profiles In Liver Cancer

: FILE REFERENCE: 44921-5028-WO

: CURRENT APPLICATION NUMBER: US/09/880,107

: PRIOR FILING DATE: 2001-06-14

: PRIOR APPLICATION NUMBER: US 60/211,379

: PRIOR FILING DATE: 2000-06-14

: PRIOR APPLICATION NUMBER: US 60/237,054

: NUMBER OF SEQ ID NOS: 3950

: SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 2370

: LENGTH: 1745

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: OTHER INFORMATION: Genbank Accession No. M61854

US-09-880-107-2370

Query Match 99.6%; Score 1739.6; DB 31; Length 1746;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ctccaatggaatccttttgggtccttgctctctctctctctctctctctctctctctcttcaa 60

Db 1 ctccaatggaatccttttgggtccttgctctctctctctctctctctctctctctctcttcaa 60

Qy 61 tctggagacagagctctgggagagaaaactcctctctctctctctctctctctctctctccagtg 120

Db 61 tctggagacagagctctgggagagaaaactcctctctctctctctctctctctctctctccagtg 120

Qy 121 ttggaaatactctcacagatagatataagagatgcagaaaactccttaaccaatctctcaa 180

Db 121 ttggaaatactctcacagatagatataagagatgcagaaaactccttaaccaatctctcaa 180

Qy 181 aaatctatggccctgtgttctcactctgtatttttggcctggaacgcatggtggtgctgcatg 240

Db 181 aaatctatggccctgtgttctcactctgtatttttggcctggaacgcatggtggtgctgcatg 240

```
QY 241 gatataagtggtgaaggagccctgattgatcttggagagagtttctggaagagcc 300
D 241 gatataagtggtgaaggagccctgattgatcttggagagagtttctggaagagcc 300
QY 301 atttccactggctgaagactaaacagagatttggaaatcttctcaacaatgaaga 360
D 301 atttccactggctgaagactaaacagagatttggaaatcttctcaacaatgaaga 360
QY 361 gatggaagagattcggcgcttctccctcatgacgctgcgggaatttgggagggga 420
D 361 gatggaagagattcggcgcttctccctcatgacgctgcgggaatttgggagggga 420
QY 421 ggagcattgagaccgtgttcaagagaaagcccgctgccttctggagaggttga 480
D 421 ggagcattgagaccgtgttcaagagaaagcccgctgccttctggagaggttga 480
QY 481 ccaaggcttccactgtatccactttcatctggctgctccctcaatgtgatct 540
D 481 ccaaggcttccactgtatccactttcatctggctgctccctcaatgtgatct 540
QY 541 gctccattttccagaagcgtttcgattataaagatcacgaatttcttaacttg 600
D 541 gctccattttccagaagcgtttcgattataaagatcacgaatttcttaacttg 600
QY 601 aaaaattgaatgaacacatcaggattgtaagcaccctcgtgacagatatacaatt 660
D 601 aaaaattgaatgaacacatcaggattgtaagcaccctcgtgacagatatacaatt 660
QY 661 ttccactatcattgattatttcccggaacccataaacaatttctaaaccttgct 720
D 661 ttccactatcattgattatttcccggaacccataaacaatttctaaaccttgct 720
QY 721 ttatgaaagtatatatttggagaaagttaaagaacaccaaagaatcgatgacata 780
D 721 ttatgaaagtatatatttggagaaagttaaagaacaccaaagaatcgatgacata 780
QY 781 acctcggaactttattgatttctctgatacaaaatgaaagaaagaaacaaacac 840
D 781 acctcggaactttattgatttctctgatacaaaatgaaagaaagaaacaaacac 840
QY 841 agtctgaattcacttgaacttggtaactcactgagctgacttacttggagctgga 900
D 841 agtctgaattcacttgaacttggtaactcactgagctgacttacttggagctgga 900
QY 901 cagagacaacagcaacccctgagatgctctctctctctctctctgagcaaccag 960
D 901 cagagacaacagcaacccctgagatgctctctctctctctctctgagcaaccag 960
QY 961 tcacagctaaagtccaggagagattgaacgtgtcattggcagaaacccggagccct 1020
D 961 tcacagctaaagtccaggagagattgaacgtgtcattggcagaaacccggagccct 1020
QY 1021 tgacagacagggccacatgcctcacagatgctgtgtgtgtgtgtgtgtgtgtgt 1080
D 1021 tgacagacagggccacatgcctcacagatgctgtgtgtgtgtgtgtgtgtgtgt 1080
QY 1081 tgacactatcccccagcctgcccctgacgtgacgtgacgtgacgtgacgtgacgt 1140
D 1081 tgacactatcccccagcctgcccctgacgtgacgtgacgtgacgtgacgtgacgt 1140
QY 1141 acctcattcccaagggaacacattataacttccctcacttctctgtcatatgaca 1200
D 1141 acctcattcccaagggaacacattataacttccctcacttctctgtcatatgaca 1200
QY 1201 aagaatttcccaaccagagatttggaccctcgtcacttctctgtgataagagga 1260
D 1201 aagaatttcccaaccagagatttggaccctcgtcacttctctgtgataagagga 1260
QY 1261 ttaagaaaagtacttacttctctctcagcaggaacacggagatttggagagag 1320
D 1261 ttaagaaaagtacttacttctctctcagcaggaacacggagatttggagagag 1320
QY 1321 gctggcccgcatggagctgttttatttctctgacatttcttcaagaacttta 1380
```

```
D 1321 gctggcccgcatggagctgttttatttcttcccgaccttatttcaagaacttta 1380
QY 1381 aatctctgattgaccccaagacattgacacaaactctgttgcataatggttct 1440
D 1381 aatctctgattgaccccaagacattgacacaaactctgttgcataatggttct 1440
QY 1441 tcccgcccttctacagctgtgcttcttctctctgctgctgaaagacacagat 1500
D 1441 tcccgcccttctacagctgtgcttcttctctctgctgctgaaagacacagat 1500
QY 1501 ctctctgctgctccctgcagctctcttctctctgctgctgctgctgctgct 1560
D 1501 ctctctgctgctccctgcagctctcttctctctgctgctgctgctgctgct 1560
QY 1561 tctctgacccgtcatctcacatttcccttccctcccaagatctagtgaaact 1620
D 1561 tctctgacccgtcatctcacatttcccttccctcccaagatctagtgaaact 1620
QY 1621 cattaaaaatttctcactgtgctgctgctgctgctgctgctgctgctgctgct 1680
D 1621 cattaaaaatttctcactgtgctgctgctgctgctgctgctgctgctgctgct 1680
QY 1681 attgagtcacacataatgctgatactgctgctgctgctgctgctgctgctgct 1740
D 1681 attgagtcacacataatgctgatactgctgctgctgctgctgctgctgctgct 1740
QY 1741 aataga 1746
D 1741 aataga 1746
```

RESULT 6

US-60-226-176-749

: Sequence 749, Application US/60226176

: GENERAL INFORMATION:

: APPLICANT: Ring, Huijun Z.

: APPLICANT: Malsen, Gareth

: APPLICANT: Townley, David

: APPLICANT: Morris, MacDonald

: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes

: FILE REFERENCE: GX-0013-1 P

: CURRENT APPLICATION NUMBER: US/60/226,176

: CURRENT FILING DATE: 2000-08-16

: NUMBER OF SEQ ID NOS: 2447

: SOFTWARE: PERL Program

: SEQ ID NO 749

: LENGTH: 1746

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: NAME/KEY: misc_feature

: OTHER INFORMATION: GB:HUMCYP219

US-60-226-176-749

Query Match

Best Local Similarity 99.8%; Score 1739.6; DB 55; Length 1746;

Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
QY 1 ctccaatggatccttttgggtgctctgtctctctctctctctctctctctctca 60
D 1 ctccaatggatccttttgggtgctctgtctctctctctctctctctctctctca 60
QY 61 tctggagacagagctctgggagagagaaactcctctgctgcccactctctccagta 120
D 61 tctggagacagagctctgggagagagaaactcctctgctgcccactctctccagta 120
QY 121 ttggaataatctcacagatagatattgaagatgcaacaatacttcaacaactctca 180
D 121 ttggaataatctcacagatagatattgaagatgcaacaatacttcaacaactctca 180
QY 181 aaatctatggccctgtgttccactctgtattttggcctggaaacgcatggtgtgcatg 240
```

Db 181 aatctatgcccctggtctcaactctgtattttggcctgaaacgcatgggtgctgcatg 240
OY 241 gatataagtggtgaaagccctgattgatcttgagagagagttttctgaaagagcc 300
Db 241 gatataagtggtgaaagccctgattgatcttgagagagagttttctgaaagagcc 300
OY 301 atttccactggtgaaagcgttaacagagagatttggaaatcgttttcagcaatgaaaga 360
Db 301 atttccactggtgaaagcgttaacagagagatttggaaatcgttttcagcaatgaaaga 360
OY 361 gatggaagagattcggtcttctccctcatgacgctgcggaattttggatggggaaga 420
Db 361 gatggaagagattcggtcttctccctcatgacgctgcggaattttggatggggaaga 420
OY 421 ggaacattgagagaccgtgttcaagagagaagccgcgtgccttgtaggagagttgagaaaaa 480
Db 421 ggaacattgagagaccgtgttcaagagagaagccgcgtgccttgtaggagagttgagaaaaa 480
OY 481 ccaaggttcacccctgattcccaactttcatctcggtgctccctgcgaatgcatct 540
Db 481 ccaaggttcacccctgattcccaactttcatctcggtgctccctgcgaatgcatct 540
OY 541 gttccatttttccagaacggttctgatataaagatcacgaattttcttaacttgatgg 600
Db 541 gttccatttttccagaacggttctgatataaagatcacgaattttcttaacttgatgg 600
OY 601 aaaaaattgaataaacaatcaggattgtaagcaccccttgatccagatgcaataatt 660
Db 601 aaaaaattgaataaacaatcaggattgtaagcaccccttgatccagatgcaataatt 660
OY 661 ttcccaactatattgattatttcccggaacccataaacaatttcttaaaaccttgctt 720
Db 661 ttcccaactatattgattatttcccggaacccataaacaatttcttaaaaccttgctt 720
OY 721 ttatggaagtgatattttggagaagtaaaagaacacagaagaatcgatgacatcaaca 780
Db 721 ttatggaagtgatattttggagaagtaaaagaacacagaagaatcgatgacatcaaca 780
OY 781 accctcggaactttattgattgcttccctgatacaaaatggagaagaaagcaaaccaac 840
Db 781 accctcggaactttattgattgcttccctgatacaaaatggagaagaaagcaaaccaac 840
OY 841 agtctgaattcactattgaaacttggtaatacactgcagctgacttacttgagctggga 900
Db 841 agtctgaattcactattgaaacttggtaatacactgcagctgacttacttgagctggga 900
OY 901 cagagacaacagcaaacccctgagatgctctctctctctctctctgctgaagcaccagagg 960
Db 901 cagagacaacagcaaacccctgagatgctctctctctctctctctgctgaagcaccagagg 960
OY 961 tcacagctaaagtcaggagaagattgaaactgtaacgtgctcattggcagaacccggagcccttga 1020
Db 961 tcacagctaaagtcaggagaagattgaaactgtaacgtgctcattggcagaacccggagcccttga 1020
OY 1021 tgcacagcaggggcccacatgcccacacagatgctgtggtgcacagagttccagagataca 1080
Db 1021 tgcacagcaggggcccacatgcccacacagatgctgtggtgcacagagttccagagataca 1080
OY 1081 tgcacacatcccacacagctgcccacatgagctgagctgagctgagctgagctgagctgagct 1140
Db 1081 tgcacacatcccacacagctgcccacatgagctgagctgagctgagctgagctgagctgagct 1140
OY 1141 acctatttcccaagggcacaacacatttacccttccctcacttctgctacatgacaaca 1200
Db 1141 acctatttcccaagggcacaacacatttacccttccctcacttctgctacatgacaaca 1200
OY 1201 aagaatttcccaacccagagatggttacccttgcacttcttctggtgagaggtggaaatt 1260
Db 1201 aagaatttcccaacccagagatggttacccttgcacttcttctggtgagaggtggaaatt 1260
OY 1261 ttaagaaaagttaactacttcatgcttcttcagcagaaacagatttggtaggagagg 1320
Db 1261 ttaagaaaagttaactacttcatgcttcttcagcagaaacagatttggtaggagagg 1320

Db 1261 ttaagaaaagttaactacttcatgcttcttcagcagaaacagatttggtaggagagg 1320
OY 1321 gcttgcccgcgatggagctgtttttattctcagccttcattttacagaactttaacctga 1380
Db 1321 gcttgcccgcgatggagctgtttttattctcagccttcattttacagaactttaacctga 1380
OY 1381 aatctctgattgacccaaagacccctgacacaaactcctgtgtcaatggatttgcttctg 1440
Db 1381 aatctctgattgacccaaagacccctgacacaaactcctgtgtcaatggatttgcttctg 1440
OY 1441 tccgcgcctttatcagctgcttctcattctcgtctctgctgaagaagcacagatggctggctg 1500
Db 1441 tccgcgcctttatcagctgcttctcattctcgtctctgctgaagaagcacagatggctggctg 1500
OY 1501 ctctctgctgctccctgcagctctcttctcctgtgtccaaatttccattctctatgatgct 1560
Db 1501 ctctctgctgctccctgcagctctcttctcctgtgtccaaatttccattctctgagct 1560
OY 1561 tctctgaccctgcatctcatttcccttcccccagatctagtgaaacatlcagcctc 1620
Db 1561 tctctgaccctgcatctcatttcccttcccccagatctagtgaaacatlcagcctc 1620
OY 1621 cattaaaaaagtttccactgtgcaaatatatactgctattcccccatactctataatagttac 1680
Db 1621 cattaaaaaagtttccactgtgcaaatatatactgctattcccccatactctataatagttac 1680
OY 1681 attgagtgccacataatgctgatacttcttaattgttgaatttataacattattatta 1740
Db 1681 attgagtgccacataatgctgatacttcttaattgttgaatttataacattattatta 1740
OY 1741 aataga 1746
Db 1741 aataga 1746

RESULT 7

US-60-233-468-749
: Sequence 749, Application US/60233468
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: APPLICANT: Valdes, Ana
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-2 P
: CURRENT APPLICATION NUMBER: US/60/233,468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 2488
: SOFTWARE: PERL Program
: SEQ ID NO 749
: LENGTH: 1746
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GB:HUMCYPC219
US-60-233-468-749

Query Match 99.6%; Score 1739.6; DB 56; Length 1746;
Best local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ctccaatggatccttttgggtccttggctctgtctctcctcatgtttgtcttcccttcaa 60
Db 1 ctccaatggatccttttgggtccttggctctgtctctcctcatgtttgtcttcccttcaa 60
OY 61 tctggagacagagctctggagaggaagaaactccctctggccccactctctccacagtga 120
Db 61 tctggagacagagctctggagaggaagaaactccctctggccccactctctccacagtga 120
OY 121 ttggaatactctacagatagattaaagatgtcagcaaaatcccttaaccaatctctcaa 180

```
Db 121 ttggaataatcctcagatagatattaggatgctcagcaaatccttaaccaatctctcaa 180
Oy 181 aaatctatgcccctgtgttcaactctgtattttggccctggaaacgcatgtgtgtgctgcatg 240
Db 181 aaatctatgcccctgtgttcaactctgtattttggccctggaaacgcatgtgtgtgctgcatg 240
Oy 241 gatataagtggtgaagaagccctgattgattctgttgagaggagtttcttggaagagcc 300
Db 241 gatataagtggtgaagaagccctgattgattctgttgagaggagtttcttggaagagcc 300
Oy 301 atttcccactggctgaaagagctaaacagagagatttggaaatcgttttccagcaatggaaaga 360
Db 301 atttcccactggctgaaagagctaaacagagagatttggaaatcgttttccagcaatggaaaga 360
Oy 361 gatagaagagatctggcgttttccctctcatgacgctgcggaaattttggatggggaaga 420
Db 361 gatagaagagatctggcgttttccctctcatgacgctgcggaaattttggatggggaaga 420
Oy 421 ggagcattggagaccgtgttcaagaggaagcccgctgccttggagagagttgaagaaaaa 480
Db 421 ggagcattggagaccgtgttcaagaggaagcccgctgccttggagagagttgaagaaaaa 480
Oy 481 ccaaggcttaacctgtgatcccaacttcatcctctgggctgtgctccctgcaatgtgatct 540
Db 481 ccaaggcttaacctgtgatcccaacttcatcctctgggctgtgctccctgcaatgtgatct 540
Oy 541 gctccattatttccagaaacgttttcgattataaagatcacgaattttttaacttgaatct 500
Db 541 gctccattatttccagaaacgttttcgattataaagatcacgaattttttaacttgaatct 500
Oy 601 aaaaattgaaatgaaacatcagagattgtgaagcaccctggatccagagatgcaataatt 560
Db 601 aaaaattgaaatgaaacatcagagattgtgaagcaccctggatccagagatgcaataatt 560
Oy 661 ttcccactatcattgattatttccgggaacccatacaaaattactttaaaaccttgctt 720
Db 661 ttcccactatcattgattatttccgggaacccatacaaaattactttaaaaccttgctt 720
Oy 721 ttatggaaatgatattttggagaagtaaaagaaaccccaagaaatcgatggacatcaaca 780
Db 721 ttatggaaatgatattttggagaagtaaaagaaaccccaagaaatcgatggacatcaaca 780
Oy 781 acctcgggactttattgattgttctctgatcaaaatggagaaggaaagcaaaaccaa 840
Db 781 acctcgggactttattgattgttctctgatcaaaatggagaaggaaagcaaaaccaa 840
Oy 841 agtctgaattcactatttgaaaacttggtaatcactgcagctgacttacttggagctggga 900
Db 841 agtctgaattcactatttgaaaacttggtaatcactgcagctgacttacttggagctggga 900
Oy 901 cagagacaacaaacacccctgagatattgtctctctctctctctctctctctctctctct 960
Db 901 cagagacaacaaacacccctgagatattgtctctctctctctctctctctctctctctct 960
Oy 961 tcacagctaaagtccaggaagatgaaacgtgtctatggcaaaacccgagccctctga 1020
Db 961 tcacagctaaagtccaggaagatgaaacgtgtctatggcaaaacccgagccctctga 1020
Oy 1021 tgacagacagggccacatgccctcacacagatgtctgtgtgacgaggttccagagatata 1080
Db 1021 tgacagacagggccacatgccctcacacagatgtctgtgtgacgaggttccagagatata 1080
Oy 1081 tcgacctcatcccccacagcctgcccctatgacgtgacctgtgacgttaaatcagaacct 1140
Db 1081 tcgacctcatcccccacagcctgcccctatgacgtgacctgtgacgttaaatcagaacct 1140
Oy 1141 acctcattcccagggaacacacatttaacttcccctcacttctgtgtgtacatgacaaca 1200
Db 1141 acctcattcccagggaacacacatttaacttcccctcacttctgtgtgtacatgacaaca 1200
Oy 1201 aagaatttcccacacagagattttgacccctgtcaccttctctctctctctctctctctct 1260
```

```
Db 1201 aagaatttcccacacacagagatgtttgacccctcgtcacttctctggatgaaggtggaatt 1260
Oy 1261 ttaagaaaagtaactacttcatgctcttctcagcagagaaaacgagatttctgtgaggagag 1320
Db 1261 ttaagaaaagtaactacttcatgctcttctcagcagagaaaacgagatttctgtgaggagag 1320
Oy 1321 gccggcccgcatggagctgtttttatttactgaccttcaatttacagaactttaacctga 1380
Db 1321 gccggcccgcatggagctgtttttatttactgaccttcaatttacagaactttaacctga 1380
Oy 1381 aatctctgattgaccccaaaagacatttgacacactcctgttgcattgattgtctctctg 1440
Db 1381 aatctctgattgaccccaaaagacatttgacacactcctgttgcattgattgtctctctg 1440
Oy 1441 tcccgcccttctatcagctgtgtcttcatctcctgtctgaagaagcacagatggtctggctg 1500
Db 1441 tcccgcccttctatcagctgtgtcttcatctcctgtctgaagaagcacagatggtctggctg 1500
Oy 1501 ctctctgctgtccttgcagctctcttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560
Db 1501 ctctctgctgtccttgcagctctcttctctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560
Oy 1561 tctctgacccgtcatctcacatttcccttcccacaaagatctagtgaacattcagctc 1620
Db 1561 tctctgacccgtcatctcacatttcccttcccacaaagatctagtgaacattcagctc 1620
Oy 1621 cattaaaaaagtctcactgtgcaaatatctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
Db 1621 cattaaaaaagtctcactgtgcaaatatctctgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 1680
Oy 1681 attgagtgccacataaagctgatacttcttaagtgtgagttataacatatattatta 1740
Db 1681 attgagtgccacataaagctgatacttcttaagtgtgagttataacatatattatta 1740
Oy 1741 aataga 1746
Db 1741 aataga 1746
```

RESULT 8

```
US-60-313-371-749
: Sequence 749, Application US/60313371
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun 2.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-S P
: CURRENT APPLICATION NUMBER: US/60/313.371
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 749
: LENGTH: 1746
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: GB:HUMCYP219
US-60-313-371-749
```

Query Match 99.6%; Score 1739.6; DB 64; Length 1746;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Oy 1 cttaaatgagatcctttgtgtctcttctgtctctctctctctctctctctctctctctca 60
Db 1 cttaaatgagatccttttggctctgtctctctctctctctctctctctctctctca 60
Oy 61 tctgagacagagctctctggagagagaaaactcctctctgcccactcctctccagtg 120
```


QY	121	ttgaaatctctacagatagatatgaaggaatgcagcaaatctctaaaccaatctctcaa	181
DB	122	ttgaaatctctacagatagatatgaaggaatgcagcaaatctctaaaccaatctctcaa	182
QY	181	aaatctatggccctgtgttcaactctgtattttggccctggaaacgcatcggtgtctcatg	240
DB	183	aaatctatggccctgtgttcaactctgtattttggccctggaaacgcatcggtgtctcatg	242
QY	241	gatatgaagtgtggaaggaagccctgattgatcttgagaggaagattttctggaagagcc	300
DB	243	gatatgaagtgtggaaggaagccctgattgatcttgagaggaagattttctggaagagcc	302
QY	301	atttccacttgctgaagagtaacagagagatttgaatcgttttcagcaatggaaaga	360
DB	303	atttccacttgctgaagagtaacagagagatttgaatcgttttcagcaatggaaaga	362
QY	361	gatggaagagagatccgcgctttccctcatgacgttcggaatttttggatggggaaga	420
DB	363	gatggaagagatccgcgctttccctcatgacgttcggaatttttggatggggaaga	422
QY	421	gagcaattggaagaccgttttcaagaggaagcccgctcttgaggagagtttgagaaaaa	480
DB	423	ggagcattgaggaccgttttcaagaggaagcccgctcttgaggagagtttgagaaaaa	482
QY	481	ccaaggtctcaccttgtaatcccaatttcaatctctgggctgtgctccctgcaatgtgat	540
DB	483	ccaaggtctcaccttgtaatcccaatttcaatctctgggctgtgctccctgcaatgtgat	542
QY	541	gctccattatttccagaaacgtttcgattataaagatcacgaatttcttaaccttgagg	600
DB	543	gctccattatttccagaaacgtttcgattataaagatcacgaatttcttaaccttgagg	602
QY	601	aaaaattgaatgaaacatcaggattgtgaagcccccctggatccagatgtgcaataat	660
DB	603	aaaaattgaatgaaacatcaggattgtgaagcccccctggatccagatgtgcaataat	662
QY	661	ttccacatcatcttgattatttcccggaacccatacaaaatttacttaaaaccttgctt	720
DB	663	ttccacatcatcttgattatttcccggaacccatacaaaatttacttaaaaccttgctt	722
QY	721	ttatggaagtgtattttggagaagtataagaaacccaagaatcgatggacatcaaca	780
DB	723	ttatggaagtgtattttggagaagtataagaaacccaagaatcgatggacatcaaca	782
QY	781	acctctggaccttattgattgcttccctgatcaaaatggagaagaaagcaaaaccaac	840
DB	783	acctctggaccttattgattgcttccctgatcaaaatggagaagaaagcaaaaccaac	842
QY	841	agctgtaattcaatttgaaaaacttggtaatcactgcagctgacttacttggagctggga	900
DB	843	agctgtaattcaatttgaaaaacttggtaatcactgcagctgacttacttggagctggga	902
QY	901	cagagacaacagcacaaacctgagatagcttctcttctctctgctgaagcaccagaag	960
DB	903	cagagacaacagcacaaacctgagatagcttctcttctctctgctgaagcaccagaag	962
QY	961	tcacagctaaagtcaggagagattgaaacgtgtcatttggcagaacccggagcccttgca	1020
DB	963	tcacagctaaagtcaggagagattgaaacgtgtcatttggcagaacccggagcccttgca	1022
QY	1021	tgcaagcaggggccacatgccctcacagatgtctgttggtgcacgaagttccagagataca	1080
DB	1023	tgcaagcaggggccacatgccctcacagatgtctgttggtgcacgaagttccagagataca	1082
QY	1081	tcgaactcatcccccacagccctgcccatgcaatgacgttgcacgtttaaattcagaacct	1140
DB	1083	tcgaactcatcccccacagccctgcccatgcaatgacgttgcacgtttaaattcagaacct	1142
QY	1141	acctcattcccaaggcacaaacattataacttccctcaacttctgtctcatatgacaaca	1200
DB	1143	acctcattcccaaggcacaaacattataacttccctcaacttctgtctcatatgacaaca	1202

QY	1201	aga	aatttcccaaccacagagagtggttgaccctcgctcaacttcttgatgaaggtgaaatt	1260
Db	1203	aga	aatttcccaaccacagagtggttgaccctcgctcaacttcttgatgaaggtgaaatt	1262
QY	1261	tla	agaaagtaactaactcaatgcctttctcaacagagaaacggatttgctgggagagg	1320
Db	1263	tla	agaaagtaactcaactcaatgcctttctcaacagagaaacggatttgctgggagagg	1322
QY	1321	gct	ggccgcgatggagctgtttttattctcgaccttcatttacagaaactttaacctga	1380
Db	1323	gct	ggccgcgatggagctgtttttattctcgaccttcatttacagaaactttaacctga	1382
QY	1381	aact	ctctgaatgacccaaaggaccttgacaaactcctgtgtcaatggattgtctctg	1440
Db	1383	aact	ctctgaatgacccaaaggaccttgacaaactcctgtgtcaatggattgtctctg	1442
QY	1441	tccg	cccttctacagctgtgcttcaactcctgtggaagacacagatgctgagctg	1500
Db	1443	tccg	cccttctacagctgtgcttcaactcctgtggaagacacagatgctgagctg	1502
QY	1501	ctct	gtgctgcctgcagctctcttctctgtgtggtccaaatttcaactatctgtgatgct	1560
Db	1503	ctct	gtgctgcctgcagctctcttctctgtgtggtccaaatttcaactatctgtgatgct	1562
QY	1561	tctt	cgaccctcatctcacatttcccttcccacagatctaagtgaacattcagctc	1620
Db	1563	tctt	cgaccctcatctcacatttcccttcccacagatctaagtgaacattcagctc	1622
QY	1621	catta	aaaaagtttccctgcycaaatatctctgtctattccccaactctataatagtac	1680
Db	1623	catta	aaaaagtttccctgcycaaatatctctgtctattccccaactctataatagtac	1682
QY	1681	attg	agtgccacataatgctgatactgtctaatgttgagittataacataattattatta	1740
Db	1683	attg	agtgccacataatgctgatactgtctaatgttgagittataacataattattatta	1742
QY	1741	aatga	1746	
Db	1743	aatga	1748	
RESULT 10				
US-09-488-127B-58				
; Sequence 58. Application US/09488127B				
; GENERAL INFORMATION:				
; APPLICANT: Thomann, Hans-Ulrich				
; APPLICANT: Fitzgerald, Michael S.				
; TITLE OF INVENTION: RAPID DETERMINATION OF GENE STRUCTURE				
; TITLE OF INVENTION: USING cDNA SEQUENCE				
; FILE REFERENCE: 2709, 1005-000				
; CURRENT APPLICATION NUMBER: US/09/488.127B				
; CURRENT FILING DATE: 2000-01-20				
; NUMBER OF SEQ ID NOS: 97				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 58				
; LENGTH: 1748				
; TYPE: DNA				
; ORGANISM: Homo sapiens				
US-09-488-127B-58				
Query Match 99.6%; Score 1739.6; DB 18; Length 1748;				
Best Local Similarity 99.8%; Pred. No. 0;				
Matches 1742; Conservative 0; Mismatches 4; Indels 0; Gaps				
QY	1	ctt	caatgatccttttggatcctttgtgctctgtctctcatgttttgcttcccttcaa	60
Db	3	ctt	caatggatccttttggctccttgctctgtctctcatgttttgcttcccttcaa	62
QY	61	ctg	ggagacagagctctagagagagaaaaactccctctctggcccaactctctccacgtga	120
Db	63	ctg	ggagagagagctctagagagagagaaaaactccctctctggcccaactctctccacgtga	120

```
Oy 121 ttgaaatactacagataagattgaaggtgctcagcaaatcttaaccaatctctcaa 180
Db 123 tcggaaatactacagataagattgaaggtgctcagcaaatcttaaccaatctctcaa 182
Oy 181 aaatatgcccctgtgttcaactctgtattttggcctggaaacgcatgtgtgctcatg 240
Db 183 aaatatgcccctgtgttcaactctgtattttggcctggaaacgcatgtgtgctcatg 242
Oy 241 gatatgaagtgtgaaaggaagccctgattgacatcttgagaggagtttctggaagagcc 300
Db 243 gatatgaagtgtgaaaggaagccctgattgacatcttgagaggagtttctggaagagcc 302
Oy 301 atttcccactggctgaaagagtaacagagatttgaatcgcttttcagcaatggaaaga 360
Db 303 atttcccactggctgaaagagtaacagagatttgaatcgcttttcagcaatggaaaga 362
Oy 361 gatggaaggaatttcggcgtttctccctcatgacgctgcggaattttggatgggaaga 420
Db 363 gatggaaggaatttcggcgtttctccctcatgacgctgcggaattttggatgggaaga 422
Oy 421 ggagcattgaggaacggtgttcaagagaagcccgctgctgtggaggagttgagaaaaa 480
Db 423 ggagcattgaggaacggtgttcaagagaagcccgctgctgtggaggagttgagaaaaa 482
Oy 481 ccaagggttcaccctgtgatcccaacttcatcctgggctgtgctccctgcaatgtgatct 540
Db 483 ccaagggttcaccctgtgatcccaacttcatcctgggctgtgctccctgcaatgtgatct 542
Oy 541 gctccattttccagaagcgtttcgattataaagatacgaatcttcaacttgatgg 600
Db 543 gctccattttccagaagcgtttcgattataaagatacgaatcttcaacttgatgg 602
Oy 601 aaaaattgaatgaaacacatcaggattgtaagcaccctcgatccagatatacaataatt 660
Db 603 aaaaattgaatgaaacacatcaggattgtaagcaccctcgatccagatatacaataatt 662
Oy 661 ttccacatcattgattatttcccggaacccatacaaaattacttaaaaccttgctt 720
Db 663 ttccacatcattgattatttcccggaacccatacaaaattacttaaaaccttgctt 722
Oy 721 ttatggaagtgtattttggagaagtaaaagaaacacaaagaaatcgatggacatcaaca 780
Db 723 ttatggaagtgtattttggagaagtaaaagaaacacaaagaaatcgatggacatcaaca 782
Oy 781 accctgggactttattgattgcttctctgatcaaaatggagaaggaagaaacacaaac 840
Db 783 accctgggactttattgattgcttctctgatcaaaatggagaaggaagaaacacaaac 842
Oy 841 agtctgaattcactattgaaacattgtaatacctcagctcacttacttgagctggga 900
Db 843 agtctgaattcactattgaaacattgtaatacctcagctcacttacttgagctggga 902
Oy 901 cagagacaacagcaaaccttgagatgctctctctctctctctctctgctggaagcaccagag 960
Db 903 cagagacaacagcaaaccttgagatgctctctctctctctctctctgctggaagcaccagag 962
Oy 961 tcacagctaaagtcaggaagattgaacgtgctcatttggaagaaacccggagccctgca 1020
Db 963 tcacagctaaagtcaggaagattgaacgtgctcatttggaagaaacccggagccctgca 1022
Oy 1021 tgcacagagggggccacatgccctacacagatgctgtgtgtgacagaggtccacagatataca 1080
Db 1023 tgcaggaacagggggccacatgccctacacagatgctgtgtgtgacagaggtccacagatataca 1082
Oy 1081 tgcactctatccccacagcctgccccatgcaatgcaatgacgtgacgttgaacttaattcagaact 1140
Db 1083 tgcactctatccccacagcctgccccatgcaatgcaatgacgtgacgttgaacttaattcagaact 1142
Oy 1141 acctatttccaaagggcacaacatatttaacttccctcacttctgtgtgtacatgacaaca 1200
Db 1143 acctatttccaaagggcacaacatatttaacttccctcacttctgtgtgtacatgacaaca 1202
Oy 1201 aagaatttcccaaccagagatttttgaccctcgctcaactttcttgatgaaggtggaatt 1260
```

```
Db 1203 aagaatttcccaaccagagatttttgaccctcgctcaactttcttgatgaaggtggaatt 1262
Oy 1261 ttaagaaaagttaactacttcaatgcctttctcagcaggaaaaacggatttgtgtggagagg 1320
Db 1263 ttaagaaaagttaactacttcaatgcctttctcagcaggaaaaacggatttgtgtggagagg 1322
Oy 1321 gctctgcccgaatggagctgttttttattccctgaaccttcatitttacagaaactttaacctga 1380
Db 1323 gctctgcccgaatggagctgttttttattccctgaaccttcatitttacagaaactttaacctga 1382
Oy 1381 aatctctgaatgaccccaaggacattgacacaaactccctgtgttcaatggatttgcctctg 1440
Db 1383 aatctctgaatgaccccaaggacattgacacaaactccctgtgttcaatggatttgcctctg 1442
Oy 1441 tcccgccttctatcagctgtcttcaacttcccttcccccagaatctagtgaaacattcagcctc 1500
Db 1443 tcccgccttctatcagctgtcttcaacttcccttcccccagaatctagtgaaacattcagcctc 1502
Oy 1501 ctctctgtctgctccctcagctctcttctctgtgctggtccaaatttcaactatctgtgatgct 1560
Db 1503 ctctctgtctgctccctcagctctcttctctgtgctggtccaaatttcaactatctgtgatgct 1562
Oy 1561 tctctgtgacccgtctatcacaatttcccttcccccagaatctagtgaaacattcagcctc 1620
Db 1563 tctctgtgacccgtctatcacaatttcccttcccccagaatctagtgaaacattcagcctc 1622
Oy 1621 cattaaaaagtttcaactgtgcaaatatctgtctattccccactactataatagttac 1680
Db 1623 cattaaaaagtttcaactgtgcaaatatctgtctattccccactactataatagttac 1682
Oy 1681 attgaatgccacataatgctcacttctcttaattgttaattgttgagttataacattattatta 1740
Db 1683 attgaatgccacataatgctcacttctcttaattgttaattgttgagttataacattattatta 1742
Oy 1741 aataga 1746
Db 1743 aataga 1748
```

```
RESULT 11
US-09-698-010-12366
: Sequence 12366, Application US/09698010
: GENERAL INFORMATION:
: APPLICANT: Williamson, Mark
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
: FILE REFERENCE: 1600.2029-001
: CURRENT APPLICATION NUMBER: US/09/698,010
: PRIOR FILING DATE: 2000-10-27
: PRIOR APPLICATION NUMBER: 60/162,358
: NUMBER OF SEQ ID NOS: 15684
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 12366
: LENGTH: 1940
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(1940)
: OTHER INFORMATION: n = A,T,C or G
US-09-698-010-12366
```

```
Query Match 93.1%; Score 1625.2; DB 27; Length 1940;
Best Local Similarity 96.3%; Pred. No. 0;
Matches 1691; Conservative 0; Mismatches 48; Indels 17; Gaps 2;
Oy 8 ggatccttttgggtccttgatgctctctctcatgttttcttctcttccaatctggag 67
Db 1 ggatccttttgggtccttgatgctctctctcatgttttcttctcttccaatctggag 60
```

Qy 68 acagagctctggagaggaacacccctctctggcccaactcctctccagtgattgaaa 127
Db 61 acagagctctggagaggaacacccctctctggcccaactcctctccagtgattgaaa 120
Qy 128 tatctacagatagatattaagagatgcagcaaatcctaaccaatctctcaaaaatcta 187
Db 121 tatctacagatagatattaagagatgcagcaaatcctaaccaatctctcaaaaatcta 180
Qy 188 tggccctgtgttcacctctgatttggcctggaaacgcatggtggtctgcatagatga 247
Db 181 tggccctgtgttcacctctgatttggcctggaaacgcatggtggtctgcatagatga 240
Qy 248 agtggtagaagagccctgattgatcttggagagaggtttctcggagagggccattccc 307
Db 241 agtggtagaagagccctgattgatcttggagagaggtttctcggagagggccattccc 300
Qy 308 actggctgaagagcctaacaagagatttgaatcggtttcaacaatggaagagatgaaa 367
Db 301 actggctgaagagcctaacaagagatttgaatcggtttcagcaatggaagagatgaaa 360
Qy 368 ggagattcggcgtttctccctcatgacgtgcggaattttgggatgggagagagcat 427
Db 361 ggagattcggcgtttctccctcatgacgtgcggaattttgggatgggagagagcat 420
Qy 428 tggagaccgtgttcaagagagaagcccgctgcctctgtgagagagttgagaaaaaccaaggc 487
Db 421 tggagaccgtgttcaagagagaagcccgctgcctctgtgagagagttgagaaaaaccaaggc 480
Qy 488 ttcaacctgtgatcccaacttcatctcgtggtgctgctccctgcaatgtgactgctccat 547
Db 481 ttcaacctgtgatcccaacttcatctcgtggtgctgctccctgcaatgtgactgctccat 540
Qy 548 tattttcagaagacgtttcgtattataaagatcacgaatttcttaacttgatgaaaaatt 607
Db 541 tattttcagaagacgtttcgtattataaagatcacgaatttcttaacttgatgaaaaatt 600
Qy 608 gaatgaagaacatcagagattgaagacccctggatccagatgaataattttcccac 657
Db 601 gaatgaagaacatcagagattgaagacccctggatccagatgaataattttcccac 660
Qy 668 tatcatgattatttccgggaaacccatacaaaattacttaaaacccctgtctttatgga 727
Db 661 tatcatgattatttccgggaaacccatacaaaattacttaaaacccctgtctttatgga 720
Qy 728 aagtgatatttggagaaagttaaagaacacccaagaatcgatggacatcaacaacctcg 787
Db 721 aagtgatatttggagaaagttaaagaacacccaagaatcgatggacatcaacaacctcg 780
Qy 788 ggaactttatgattgctctctgatcaaaatggagaagaaaaacaaacacagctctga 847
Db 781 ggaactttatgattgctctctgatcaaaatggagaagaaaaacaaacacagctctga 840
Qy 848 attcaactattgaaacacttgtaatacactgcagctgacttacttggagctgggacagagc 907
Db 841 attcaactattgaaacacttgtaatacactgcagctgacttacttggagctgggacagagc 900
Qy 908 aacaagcacacccctgagatagctctctctctctctctctctctctctctctctctcagc 967
Db 901 aacaagcacacccctgagatagctctctctctctctctctctctctctctctctctcagc 960
Qy 968 taaagtcaggagaagatgaacgtgtcatttggcagaaacccggagcccttgatgacga 1027
Db 961 taaagtcaggagaagatgaacgtgtcatttggcagaaacccggagcccttgatgacga 1020
Qy 1028 caggggccacatgccctacacagatgctgtgtgtgcagaggtccagagatacatcgacct 1087
Db 1021 caggggccacatgccctacacagatgctgtgtgtgcagaggtccagagatacatcgacct 1080
Qy 1088 catccccacagcctgcccacagtgacctgtgacgttaaatcagaaactacatccat 1147
Db 1081 catccccacagcctgcccacagtgacctgtgacgttaaatcagaaactacatccat 1140

Qy 1148 tcccaagggcacaaccataattaacttccctcacttctgtgtctacatgacacaaagaatt 1207
Db 1141 tcccaagggcacaaccataattaacttccctcacttctgtgtctacatgacacaaagaatt 1200
Qy 1208 tcccaaccagagatgtttgacccctgcacatttctggatgaaggtggaaattttaagaa 1267
Db 1201 tcccaaccagagatgtttgacccctgcacatttctggatgaaggtggaaattttaagaa 1260
Qy 1268 aagtaactacttcacgcttctcagcaggaacacgatttctgtggagagggccctggc 1327
Db 1261 aagtaactacttcacgcttctcagcaggaacacgatttctgtggagagggccctggc 1320
Qy 1328 ccgcattggagctgtttttattctcagccttcattttacagaaactttaacctgaactctct 1387
Db 1321 ccgcattggagctgtttttattctcagccttcattttacagaaactttaacctgaactctct 1380
Qy 1388 gattgaccacaaagacactgcacacactcctgtgtcaatggatttcttctgtcccgcc 1447
Db 1381 ggttgaccacaaagacactgcacacactccagttgtcaatggatttcttctgtcccgcc 1440
Qy 1448 ctctcatcagctgtcttctcattcctgtctgaagaacacagatggtctggctgctcctgt 1507
Db 1441 ctctcatcagctgtcttctcattcctgtctgaagaacacagatggtctggctgctcctgt 1500
Qy 1508 gctgctccctcagctcttcttctctctg -----tccaaatttcaactatctgtgct 1560
Db 1501 gcagtccctcagctcttcttctctctctctctctctctctctctctctctctctctg 1560
Qy 1561 tcttgacacccgtcatctcacatttctctccccaagaatctagtgaaacattcagcctc 1620
Db 1561 tcttccacctgcacatcacatttctctctctctctctctctctctctctctctctctc 1620
Qy 1621 catta-----aaaaagtttcaactgtgcaaatatctctctctctctctctctctc 1670
Db 1621 cattcagagaggtttctctatgttctcactgtgcaaatatctctctctctctctctctg 1680
Qy 1671 taatagttacattgagtgccacataatgctgatactgtcttaattgttgatttaacat 1730
Db 1681 taacagttgcatgactgctcaataatgctcactatctataatgttgatttaatat 1740
Qy 1731 attattattataataga 1746
Db 1741 gttattattataataga 1756

RESULT 12
PCT-US98-16979-4
: Sequence 4, Application PC/TUS9816979
: GENERAL INFORMATION:
: APPLICANT: University of Rochester
: TITLE OF INVENTION: FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA
: FILE REFERENCE: 176/60232
: CURRENT APPLICATION NUMBER: PCT/US98/16979
: CURRENT FILING DATE: 1998-08-17
: EARLIER APPLICATION NUMBER: 60/056,754
: EARLIER FILING DATE: 1997-08-20
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1845
: TYPE: DNA
: ORGANISM: mammalian
PCT-US98-16979-4

Query Match 87.9%; Score 1534.6; DB 1: Length 1845;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps 2:
Qy 1 ctcaatggatccttttggctctgtgctctctctctctctctctctctctctctctca 60
Db 5 ctcaatggatccttctgtggtccttctgtctctctctctctctctctctctctctc 65

Oy 61 tctggagacagagctctgggagaggaagaaactccctctctctgcccactctctctccagtgga 120
Db 66 tctggagacagagctctgggagaggaagaaactccctctctctgcccactctctctccagtgga 125
Oy 121 ttggaataatcctcacagatagattaaagagatcgacaataccttaaccaatctctcaaa 180
Db 126 ttggaataatcctcacagatagattaaagagatcgacaataccttaaccaatctctcaaa 185
Oy 181 aaatctatgcccctctgttctcaactctgtattttggcctggaacatcggtggtgctcatg 240
Db 186 aggtctatgcccctctgttctcaactctgtattttggcctggaacatcggtggtgctcatg 245
Oy 241 gatatgaatggtggaaggaagccctgattgatatctggagagagagttttctggaagagcc 300
Db 246 gatatgaagcagtggaaggaagccctgattgatatctggagagagagttttctggaagagga 305
Oy 301 atttcccactggctggaagagcttaacagagagatttgaatcggtttctcagcaatggaaga 360
Db 306 ttttcccactggctggaagagcttaacagagagatttgaatcggtttctcagcaatggaaga 365
Oy 361 gatggaagggatttcggcgtttctccctcatgacgctgcggaattttgggatgggaaga 420
Db 366 aatggaagggatttcggcgtttctccctcatgacgctgcggaattttgggatgggaaga 425
Oy 421 ggagcattggagccgtgttcaagaggaagacccgctgcctgtgagagagttgagaaaaa 480
Db 426 ggagcattggagccgtgttcaagaggaagacccgctgcctgtgagagagttgagaaaaa 485
Oy 481 ccaaggttccacctgtgacctctctccactttcatctgggctgtgctccctgcgaatgcatct 540
Db 486 ccaaggttccacctgtgacctctctccactttcatctgggctgtgctccctgcgaatgcatct 545
Oy 541 gctccatttttccagaaagcttctcattataaagatcacgaatttcttaacttgatgg 600
Db 546 gctccatttttccagaaagcttctcattataaagatcacgaatttcttaacttgatgg 605
Oy 601 aaaaattgaataaaacatcaggatttgaagacccctctggatccagatgcaataatt 660
Db 606 aaaaattgaataaaacatcaggatttgaagacccctctggatccagatgcaataatt 665
Oy 661 ttccactatcattgatttttccgggaacccatacaaatcttaaaaacctgtctt 720
Db 666 ttctctcatcattgatttttccgggaacccatacaaatcttaaaaacctgtgtt 725
Oy 721 ttatggaagtgatttttggagaagtaaaagaaacacccaagaatcgatggacatcaaca 780
Db 726 ttatggaagtgatttttggagaagtaaaagaaacacccaagaatcgatggacatgaaca 785
Oy 781 accctcggacctttattgattgctctctgatcaaaaatggagaagaaacacaaacacac 840
Db 786 accctcggacctttattgattgctctctgatcaaaaatggagaagaaacacaaacacac 845
Oy 841 agtctgaattcactattgaaaacttggttaactcactgcagctgacttacttggagctggga 900
Db 846 catctgaatttactattgaaagcttggaacacactgcagttgacttgtttggagctggga 905
Oy 901 cagagacaacagacacacctgagatgctctctctctctctctactaaagacacacagag 960
Db 906 cagagacacagacacacctgagatgctctctctctctctctactaaagacacacagag 965
Oy 961 tcacagcttaaaagtcagagaagattgaacgtgtcatttggcagaacacggagcccttga 1020
Db 966 tcacagcttaaaagtcagagaagattgaacgtgtcatttggcagaacacggagcccttga 1025
Oy 1021 tgacagacaggggacacatgcctcacagatgctgtggtgcagaggtccagagataca 1080
Db 1026 tgacagacaggggacacatgcctcacagatgctgtggtgcagaggtccagagataca 1085
Oy 1081 tcgacctatccccaccagctgccccatgcagtgacctgtgacgtttaaatttcagaaact 1140
Db 1086 ttgaccttctccccaccagctgccccatgcagtgacctgtgacattaaatttcagaaact 1145
Oy 1141 acctcattccccagggacaaacattataacttccctcacttctgtgctacatgaaca 1200

Db 1145 atctcattccccagggcacaaacattataatttccctgactctgtgctacatgaaca 1205
Oy 1201 aagaattttcccaacccagagatgtttgacctcgtcactttcttgatgaagtggaact 1260
Db 1206 aagaattttcccaacccagagatgtttgacctcgtcactttcttgatgaagtggaact 1265
Oy 1261 ttaaaaaagtaactactactcattcattcctcagcaggaacacggatttgtgtggagagg 1320
Db 1266 ttaaaaaagtaactactactcattcattcctcagcaggaacacggatttgtgtggagagg 1325
Oy 1321 gctcggccgcagtgagctgtttttattcctgaccttattttacagaactttaacctga 1380
Db 1326 cctcggccgcagtgagctgtttttattcctgaccttattttacagaactttaacctga 1385
Oy 1381 aatctctgattgacccaaaggaccttgacaaactcctctgttcaalgatgttctctg. 1440
Db 1386 aatctctggttgacccaaaggaccttgacaaactcctctgttcaalgatgttctctg 1445
Oy 1441 tcccgccctctatcagctgtcttctcattcctctgctgaagaagacagatggtctgctg 1500
Db 1446 tgcggccctctaccagctgtgcttctcattcctctgctgaagaagacagatggcctgctg 1505
Oy 1501 ctctctgctctcctgcagctctcttctctctg-----tccaaatttcaactatctg 1553
Db 1506 ctgctgctcagctctcctgctctcttctctctggtgggacattatccactcttcaactatctg 1565
Oy 1554 tgaagcttctctgacctcctcattcctcatttctccttccccaagatctagtgaaact 1613
Db 1566 taatgcttcttccactctcattcctcatttctccttccccaagatctagtgaaact 1625
Oy 1614 cagctccattt-----aaaaagtttcaactgtcaaatatctctgtctattcccc 1663
Db 1626 cgacctcatttgcagagagtttctctgttctcactgtgcaaatatctctgtctattcccc 1685
Oy 1664 tactctataatgattcagtgacacataatgctgatactgtctaatgttgagtt 1723
Db 1686 tactctgtaacagttcagtgactgctcacaataatgctcactatctactatgttgagtt 1745
Oy 1724 ttaacattattattataaataga 1746
Db 1746 ttaattgttattattataaataga 1768

RESULT 13

US-09-023-655-1057
: Sequence 1057, Application US/09023655
: GENERAL INFORMATION:
: APPLICANT: Cocks, Benjamin G.
: APPLICANT: Susan G. Stuart
: APPLICANT: Jeffrey J. Seilhamer
: TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
: TITLE OF INVENTION: EXPRESSION
: NUMBER OF SEQUENCES: 1508
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
: CURRENT APPLICATION DATA: US/09/023, 655
: FILING DATE: HEREWITH
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:


```

US-09-135-296-4
: Sequence 4, Application US/09135296
: GENERAL INFORMATION:
: APPLICANT: Jones, Jeffrey P.
: APPLICANT: Shimoji, Miyuki
: TITLE OF INVENTION: FUNCTIONAL BACTERIAL/MAMMALIAN CYTOCHROME P450 CHIMERA
: FILE REFERENCE: 176/60231
: CURRENT APPLICATION NUMBER: US/09/135,296
: EARLIER FILING DATE: 1998-08-17
: EARLIER APPLICATION NUMBER: 60/056,754
: EARLIER FILING DATE: 1997-08-20
: NUMBER OF SEQ ID NOS: 14
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 1845
: TYPE: DNA
: ORGANISM: mammalian
US-09-135-296-4

Query Match      87.9%; Score 1534.6; DB 15; Length 1845;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps

Qy 1 ctccaatggatcccttttgggtcccttgctctgctctctcatgtttgtgtctctcttca 50
   |||||
Db 6 ctccaatggatctcttgggtcccttgctctgctctcatgtttgtgtctcttcttca 55
   |||||

Qy 61 tctggagacagaagctctgggagagaaactccctctctggcccaactctctcccagtg 120
   |||||
Db 66 tctggagacagaagctctgggagagaaactccctctctggcccaactctctcccagtg 125
   |||||

Qy 121 ttggaatatcctacagatagattaaaggatgctagcaaatccttaaccaatctctca 180
   |||||
Db 126 ttggaatatcctacagataggtattgaggacatcagcaaatccttaaccaatctctca 185
   |||||

Qy 181 aaatctagccctgtttcaactctgtattttggcctggaacgcatggtggtgctgcagt 240
   |||||
Db 186 aggtctatggcctgtttcaactctgtattttggcctgaaaccatagtggtgctgcagt 245
   |||||

Qy 241 gatataaagtgtggaagaaacccctgattgattctggagagaggagttcttggaaaggcc 300
   |||||
Db 246 gatataaagcagtgaagaaacccctgattgattctggagagaggagttcttggaaaggcc 305
   |||||

Qy 301 atttcccactgctgaagagactaacagagagatttggaaatcgtttcagcaatggaaaga 360
   |||||
Db 306 tttcccactggctgaagagactaacagagagatttggaaatcgtttcagcaatggaaaga 365
   |||||

Qy 361 gatgaagagagattcggcgcttctccctcatgacgctgcggaatttggatgggaaaga 420
   |||||
Db 366 aatggaagagagatccggcgcttctccctcatgacgctgcggaatttggatgggaaaga 425
   |||||

Qy 421 gaagcattgagaccgctgttcaagagaagaccgcgtgccttggaggagttgagaaaa 480
   |||||
Db 426 ggagcattgagaccgctgttcaagagaagaccgcgtgccttggaggagttgagaaaa 485
   |||||

Qy 481 ccaaggcttcacctgtgatcccaacttcatcctggcgtgctccctgcaatgtgatct 540
   |||||
Db 486 ccaaggcttcacctgtgatcccaacttcatcctggcgtgctccctgcaatgtgatct 545
   |||||

Qy 541 gctccattatttccagaagaacgtttcgattataaagatcacgaatttcttaacttgatgg 600
   |||||
Db 546 gctccattatttccataaaacgttttgattataaagatcagcaatttcttaacttaatgg 605
   |||||

Qy 601 aaaaattgaatgaacaatcagatgtgaagacccttgatccagatataataatt 660
   |||||
Db 606 aaaaattgaatgaacaatcagatgttgaagacccttgatccagatataataatt 665
   |||||

Qy 661 ttcccactatcattgattatttccggggaaccataaacaattacttaaaaccttgct 720
   |||||
Db 666 ttctctctatcattgattatttccggggaaccataaacaattacttaaaaccttgct 725
   |||||

Qy 721 ttatggaaagtgatttttggagaagtgataaagaacaccacaagaatctgattggacataca 780
   |||||

```


Search completed: April 19, 2002, 11:02:26
Job time: 13246 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 11:06:02 : Search time 232.27 Seconds
(without alignments)
376.816 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcagaaattttgaagcctg.....ttagctcatgtgaagcggg ~43

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 78474 seqs, 58898373 residues

Total number of hits satisfying chosen parameters: 155948

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending_Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/US05_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US05_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	40.4	5.4	13327	6	US-10-105-299-7201
C 2	35	4.7	11146	6	US-10-105-299-12104
C 3	34.6	4.7	3220	6	US-10-103-313-100
C 4	33.8	4.5	256	5	US-09-975-254-20507
C 5	33.2	4.5	52845	6	US-10-105-299-12317
C 6	33	4.4	34762	6	US-10-105-299-11849
C 7	32.8	4.4	272	6	US-10-102-524-480
C 8	32.2	4.3	261	5	US-09-975-254-24327
C 9	31.8	4.3	68571	7	US-60-368-184-1
C 10	31.6	4.3	555	6	US-10-106-698-785
C 11	31	4.2	265	5	US-09-789-189-1504
C 12	31	4.2	6833	6	US-10-105-299-7023
C 13	30.8	4.1	658	5	US-10-102-805-359
C 14	30.8	4.1	748	6	US-10-106-698-1347
C 15	30.6	4.1	49375	6	US-10-105-299-12428
C 16	30.4	4.1	671	6	US-10-105-698-1785
C 17	29.8	4.0	1134	5	US-10-105-299-1744
C 18	29.8	4.0	2007	6	US-10-106-698-1481
C 19	29.8	4.0	2893	5	US-10-105-299-11321
C 20	29.8	4.0	31304	6	US-10-105-299-12035
C 21	29.6	4.0	271	5	US-09-975-254-11859
C 22	29.6	4.0	51961	6	US-10-105-299-6419
C 23	29.4	4.0	1306	6	US-10-105-299-14391
C 24	29.4	4.0	1325	6	US-10-105-299-2832
C 25	29.4	4.0	2037	5	US-09-978-403A-429
C 26	29.4	4.0	2037	5	US-09-978-544A-429

C 27 29.4 4.0 2037 5 US-09-978-581A-429 Sequence 429, App
C 28 29.4 4.0 2037 5 US-09-978-757A-429 Sequence 429, App
C 29 29.4 4.0 2037 5 US-09-978-564A-429 Sequence 429, App
C 30 29.4 4.0 2037 5 US-09-999-831A-429 Sequence 429, App
C 31 29.4 4.0 2037 5 US-09-999-829A-429 Sequence 429, App
C 32 29.4 4.0 2037 5 US-10-013-921A-429 Sequence 429, App
C 33 29.4 4.0 2037 6 US-10-013-929A-429 Sequence 429, App
C 34 29.4 4.0 2037 6 US-10-013-918A-429 Sequence 429, App
C 35 29.4 4.0 2037 5 US-10-017-082A-429 Sequence 2363, App
C 36 29.4 4.0 2037 6 US-10-105-299-2363 Sequence 12135, A
C 37 29.4 4.0 10478 6 US-10-105-299-12136 Sequence 12135, A
C 38 29.4 4.0 14550 6 US-10-105-299-12135 Sequence 12432, A
C 39 29.4 4.0 42299 6 US-10-105-299-12432 Sequence 9312, App
C 40 29.2 3.9 23885 6 US-10-105-299-9912 Sequence 1811, App
C 41 29 3.9 351 5 US-09-789-189-1811 Sequence 91, Appl
C 42 29 3.9 9323 7 US-60-365-264-91 Sequence 6774, App
C 43 29 3.9 10554 6 US-10-105-299-6774 Sequence 6773, App
C 44 29 3.9 10558 6 US-10-105-299-6773 Sequence 1057, App
C 45 28.8 3.9 1426 6 US-10-105-299-1057 Sequence 1057, App

ALIGNMENTS

RESULT 1
US-10-105-299-7201/c
: Sequence 7201, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: ROSEN, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7201
: LENGTH: 13327
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-7201

Query Match 5.4%; Score 40.4; DB 6; Length 13327;
Best Local Similarity 53.9%; Pred. No. 0.031;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 453 aggaattgttttcagcaatgaagaagaatgaagagatccggcgttttccctcatgac 522
Db 10538 AGGAATCTTGTTTCTAACCAGCCGCTGCTGGTGACACTGCGCAATTTTGCACTTGGAGC 10479
Qy 523 gctcgggaatttggatggagagagacgattgagaccgtgttcaagagagacccg 582
Db 10478 GCTTANGAGTTCGGGTTCGGTACCGGACCGCTCGAGGCGCGCTCTTGGAGAGCGCGC 10419
Qy 583 ctgccttgagagagattgagaaaaaaccaagggt 616
Db 10418 TIGCTGCIAGACGAATTTCAAGCCACCATTCCT 10385
RESULT 2
US-10-105-299-12104/c
: Sequence 12104, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: ROSEN, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 12104

Db 1071 AGICAGGGTGACGA 1057

RESULT 10

US-10-106-698-786
: Sequence 786, Application US/10106598
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
: FILE REFERENCE: PA005PI
: CURRENT APPLICATION NUMBER: US/10/106,698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCI/US00/25524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO 785
: LENGTH: 555
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-106-698-786

Query Match 4.3%; Score 31.6; DB 6; Length 555;
Best Local Similarity 58.5%; Pred. No. 2.7;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Oy 21 tctggctgaataaaagcctacatacaatacaataatcatcgtcaataatcaggcttagca: 80
Db 80 tttctcgtggaatgacataaaataaaataaaagactttttbaaaatgaatgcttga: 139
Oy 81 atgagcaaaatgaacttcgtttgctgtttatct 114
Db 140 attgataaaactgtcttctctttttttct 173

RESULT 11

US-09-789-189-1604/c
: Sequence 1604, Application US/09789189
: GENERAL INFORMATION:
: APPLICANT: Lellias, Jean-Michel
: TITLE OF INVENTION: Human Polynucleotides and Polypeptides
: FILE REFERENCE: 25436/1720
: CURRENT APPLICATION NUMBER: US/09/789,189
: PRIOR FILING DATE: 2001-02-20
: PRIOR APPLICATION NUMBER: 60/183452
: PRIOR FILING DATE: 2000-02-18
: NUMBER OF SEQ ID NOS: 2005
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 1604
: LENGTH: 266
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-789-189-1604

Query Match 4.2%; Score 31; DB 5; Length 266;
Best Local Similarity 59.8%; Pred. No. 2.8;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Oy 5 aaatattgaagcctgtgtgctgaataaaagcctacatacaatacaaa: gaaatattcatgaa: 4;
Db 125 AAATATGGTAAGATATTGTATTATTAAGGAATTAAGGAATTAATTTTAA: 46
Oy 65 aaatcaggcttagcaaatgacaaaat 91
Db 65 GATAAAGGTTTAGGAAAGTAGAAAT 39

RESULT 12

US-10-105-299-7023
: Sequence 7023, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Rosen, et al.
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-26
: NUMBER OF SEQ ID NOS: 15197
: PRIOR APPLICATION removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7023
: LENGTH: 5833
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-7023

Query Match 4.2%; Score 31; DB 6; Length 6833;
Best Local Similarity 45.2%; Pred. No. 15;
Matches 103; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Oy 73 cttagcaaatggacacaaatagtaacttcgttgcgttattctctgtcttcttcttagct 132
Db 5471 ctgtcgttggaacagagagagaacaattcctctctcatcctaagtgtaatttggacct 553
Oy 133 ctcaaaagcttatggcctgtgttcactctgtattttggcctgaaacccatagtgtgt 192
Db 5531 atttctatccttcgcccataccctcgtctcccaatctctgaaacccactgttatgt 559
Oy 193 gcattgatagaacagcagtgaggagagccctgattgatcttggaggagagtttctgaaag 252
Db 5591 ctctagatctattgcaaaagacctttatttgggtctacatgagtgagatgtggcaa 565
Oy 253 agcatttccactgcctgaggaagactaacagagagatttgggt 295
Db 5651 tgttttcttctacctgtttagctatttaccatgcggt 5693

RESULT 13

US-10-102-806-359/c
: Sequence 359, Application US/10102806
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA103PIC1
: CURRENT APPLICATION NUMBER: US/10/102,806
: CURRENT FILING DATE: 2002-03-22
: PRIOR APPLICATION NUMBER: 09/925,298
: PRIOR FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCI/US00/05881
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 846
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 359
: LENGTH: 668
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (15)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (19)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (20)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (255)

: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (512)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (552)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (558)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (559)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (579)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (588)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (593)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (599)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc_feature
: LOCATION: (567)
: OTHER INFORMATION: n equals a.t.g. or c
US-10-102-806-359

Query Match 4.1% Score 30.8; DB 6; Length 452;
Best Local Similarity 51.2%; Pred. No. 5.2;
Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
Qy 32 aaaaagcacaatacaataaataatcatgctaaatcaggcttagcaaaatggacaaaat 31
Db 586 AAAAAAANCTAGAAAATTAAGAAAGNNCTCCANATACCCCATTTCCATAGCAATTAACA 527
Qy 92 agtaactcgtttgctgttctctgtctacttctcctagctctcaaggctctatggccct 151
Db 525 AGCTAGTATATATGTTGGTGTGTATGTATGTATGTATGTATGTATGTATGTATGTAT 457
Qy 152 gtgtcca 158
Db 466 TTTTITA 450

RESULT 14
US-10-106-698-1347/c
: Sequence 1347, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polypeptides and Polypeptide
: FILE REFERENCE: PA005PI
: CURRENT APPLICATION NUMBER: US/10/106,698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO 1347
: LENGTH: 748
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (675)..(676)
: OTHER INFORMATION: n equals a.t.g. or c

: NAME/KEY: misc_feature
: LOCATION: (734)..(734)
: OTHER INFORMATION: n equals a.t.g. or c
US-10-106-598-1347

Query Match 4.1% Score 30.8; DB 6; Length 748;
Best Local Similarity 50.7%; Pred. No. 5.5;
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
Qy 150 ctgtattcactctctattttggcctgaaccatagtgctgctgcatgataaagcag 209
Db 409 CTGTCTGACTGTGACTGATTAGGAGACACTGTTAAGGAGAGGGATGACTCTGAATCG 350
Qy 210 taaaggaagccctgattatctgttgaggagagagatlltclggaagagagcatlccactgg 269
Db 349 TGAACAATGCTGTGACTGGGCATTACTGGAGCTGGAGAGGCCATGGTTCTGGAGTGG 290
Qy 270 ctgaaagagactaacagagagatttgt 295
Db 289 TATACAGTGTGGGATGGGGTGGGT 264

RESULT 15
US-10-105-299-12428/c
: Sequence 12428, Application US/10105299
: GENERAL INFORMATION:
: APPLICANT: Fosen, et. al
: TITLE OF INVENTION: Human Secreted Proteins
: FILE REFERENCE: PS950
: CURRENT APPLICATION NUMBER: US/10/105,299
: CURRENT FILING DATE: 2002-03-25
: NUMBER OF SEQ ID NOS: 15197
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 12428
: LENGTH: 49375
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-105-299-12428

Query Match 4.1% Score 30.6; DB 6; Length 49375;
Best Local Similarity 52.8%; Pred. No. 53;
Matches 66; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 11 ttgaagcctgtgtgctgaataaaagcacaatacaatacaataatgataatcaatca 70
Db 41388 TTTAAATTTTATTTTCAGGCATACAATAATATAGATGCAATAATATACAAAATCA 41329
Qy 71 ggcttagcaaatggacaaaatagtaacttctgttgcgttatctctgtacttctctag 130
Db 41328 AGCTTAAATAATAATATGATAGATAGTACTACTGAAGCCTCGTTACCTCTTAGATCCTAT 41269
Qy 131 ctctc 135
Db 41268 TTTCC 41264

Search completed: April 19, 2002, 11:06:57
Job time: 10737 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 11:02:26 ; Search time 5830.82 Seconds
(without alignments)
2760.419 Million cell updates/sec

Title: US-09-763-292-2

Perfect score: 743

Sequence: 1 tcagaaatattgaagcctg.....ttagctcatgtgaagcgagg 743

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues

Total number of hits satisfying chosen parameters: 44046606

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA_Main:*

1: /cgn2_6/ptodata/2/pna/US0611_COMB.seq:*

2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*

3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*

4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*

5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*

6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*

7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*

8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*

9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*

10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*

11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*

12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*

13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*

14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*

15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*

16: /cgn2_6/ptodata/2/pna/US092_COMB.seq:*

17: /cgn2_6/ptodata/2/pna/US093_COMB.seq:*

18: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*

19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*

20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*

21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*

22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*

23: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*

24: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*

25: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*

26: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*

27: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*

28: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*

29: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*

30: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*

31: /cgn2_6/ptodata/2/pna/US098_COMB.seq:*

32: /cgn2_6/ptodata/2/pna/US099_COMB.seq:*

33: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*

34: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*

35: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*

36: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*

37: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*

38: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*

39: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*

40: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*

41: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*

42: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*

43: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*

44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*

45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*

46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*

47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*

48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*

49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*

50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*

51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*

52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*

53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*

54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*

55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*

56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*

57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*

58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*

59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*

60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*

61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*

62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*

63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*

64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*

65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*

66: /cgn2_6/ptodata/2/pna/US60_MERGED_COMB.seq:*

67: /cgn2_6/ptodata/2/pna/US10_MERGED_COMB.seq:*

68: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

69: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

70: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

71: /cgn2_6/ptodata/2/pna/US09_MERGED_COMB.seq:*

72: /cgn2_6/ptodata/2/pna/US08_MERGED_COMB.seq:*

73: /cgn2_6/ptodata/2/pna/US07_MERGED_COMB.seq:*

74: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq:*

75: /cgn2_6/ptodata/2/pna/US06_MERGED_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	743	100.0	743	30	US-09-763-292-2	Sequence 2, Appli
2	739.8	99.6	743	55	US-60-226-176-756	Sequence 756, App
3	739.8	99.6	743	56	US-60-233-458-756	Sequence 756, App
4	739.8	99.6	743	64	US-60-313-371-756	Sequence 15, Appl
5	728.8	98.1	7609	54	US-60-212-657-15	Sequence 56, Appl
6	728.8	98.1	32768	54	US-60-213-181-56	Sequence 24, Appl
C 7	708	95.3	51955	57	US-60-248-498-24	Sequence 61, Appl
C 8	708	95.3	222537	57	US-60-248-542-62	Sequence 62, Appl
C 9	708	95.3	222537	57	US-60-248-542-62	Sequence 9, Appl
C 10	708	95.3	225532	57	US-60-245-227-9	Sequence 59, Appl
11	622.6	83.8	8437	18	US-09-488-127-59	Sequence 59, Appl
12	622.6	83.8	8437	18	US-09-488-127B-59	Sequence 97, Appl
13	622.6	83.8	8437	18	US-09-488-127B-97	Sequence 147, App
14	555.8	74.8	572	51	US-60-182-895-147	Sequence 148, App
15	555.8	74.8	572	51	US-60-182-895-148	Sequence 149, App
16	555.8	74.8	572	51	US-60-182-895-149	Sequence 150, App
17	555.8	74.8	572	51	US-60-182-895-150	Sequence 151, App
18	555.8	74.8	572	51	US-60-182-895-151	Sequence 152, App
19	555.8	74.8	572	51	US-60-182-895-152	Sequence 18, Appl
20	555.8	74.8	572	51	US-60-182-895-152	Sequence 31, Appl
21	544.2	73.2	1580	54	US-60-213-795-31	Sequence 46, Appl
22	544.2	73.2	1580	54	US-60-213-795-31	Sequence 108, App
C 23	532.4	71.7	722	52	US-60-195-048-108	Sequence 109, App
C 24	532.4	71.7	722	52	US-60-195-048-109	Sequence 3, Appli
C 25	521.2	70.1	710	53	US-60-205-169-3	Sequence 178, App
26	488.2	65.7	664	52	US-60-198-818-178	Sequence 49, Appl
27	433.6	56.4	1001	26	US-09-671-317-49	Sequence 5, Appli
C 28	433.6	56.4	32141	55	US-60-229-511-5	Sequence 23, Appl
29	433.6	58.4	36191	57	US-60-248-498-23	Sequence 10, Appl
30	433.6	58.4	36584	57	US-60-245-227-10	Sequence 59, Appl
31	433.6	58.4	118047	57	US-60-248-498-59	

32 433.6 58.4 118047 57 US-60-248-542-60 Sequence 60, Appl
33 421.8 56.8 1001 26 US-09-671-317-50 Sequence 50, Appl
34 415.4 55.9 436 26 US-09-671-317-55 Sequence 55, Appl
c 35 369 49.7 734 18 US-09-488-127-64 Sequence 64, Appl
c 36 369 49.7 734 18 US-09-488-127B-64 Sequence 64, Appl
37 366.4 49.3 605 53 US-60-205-169-6 Sequence 6, Appl
38 366.2 49.3 557 18 US-09-488-127-62 Sequence 52, Appl
39 366.2 49.3 557 18 US-09-488-127B-62 Sequence 52, Appl
c 40 346.4 46.6 16238 54 US-60-212-657-14 Sequence 14, Appl
41 331.2 44.6 654 55 US-60-226-176-740 Sequence 740, App
42 331.2 44.6 654 56 US-60-233-468-740 Sequence 740, App
43 331.2 44.6 654 64 US-60-313-371-740 Sequence 740, App
c 44 325 43.7 1595 53 US-60-200-392-43 Sequence 43, Appl
c 45 325 43.7 1595 53 US-60-200-392-44 Sequence 44, Appl

ALIGNMENTS

RESULT 1
US-09-763-292-2
: Sequence 2, Application US/09763292
: GENERAL INFORMATION:
: APPLICANT: AB Sangtec Medical
: TITLE OF INVENTION: New method
: FILE REFERENCE: primers
: CURRENT APPLICATION NUMBER: US/09/763.292
: CURRENT FILING DATE: 2000-02-21
: NUMBER OF SEQ ID NOS: 19
: SEQ ID NO 2
: LENGTH: 743
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-763-292-2

Query Match 100.0%; Score 743; DB 30; Length 743;
Best Local Similarity 100.0%; Pred. No. 3.7e-210;
Matches 743; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 tcagaaatatttgagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Db 1 tcagaaatatttgagcctgtgtgctgaataaaagcacaatacaatacaataatca 60
Qy 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgtgttctgtct 120
Db 61 tgcataatcaggcttagcaaatgacaaaatagtaacttcgtttgtgttctgtct 120
Qy 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180
Db 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaacc 180
Qy 181 catagtgtgctgatgatgaatgaagcagtgaaaggaagccctgattgatcttgagagaga 240
Db 181 catagtgtgctgatgatgaatgaagcagtgaaaggaagccctgattgatcttgagagaga 240
Qy 241 gttttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttgtaggtg 300
Db 241 gttttctggaagaggcattttcccaactggctgaaagagctaaacagagagatttgtaggtg 300
Qy 301 tgcattgctgtttccagcattctcttggggaatggggagatgggaaacagagagacttac 360
Db 301 tgcattgctgtttccagcattctcttggggaatggggagatgggaaacagagagacttac 360
Qy 361 agagctctcgggcagagcttgcccatccacatggctgcccagtgcaacttctcttt 420
Db 361 agagctctcgggcagagcttgcccatccacatggctgcccagtgcaacttctcttt 420
Qy 421 ctgctcgggagatctccctcctagtttcttctctctctctggttaggaattgttttcagcaa 480
Db 421 ctgctcgggagatctccctcctagtttcttctctctctctggttaggaattgttttcagcaa 480

Qy 481 tggaaagaaatggagagagatccggcgttttccctcatgacgctgcggaattttgggat 540
Db 481 tggaaagaaatggagagagatccggcgttttccctcatgacgctgcggaattttgggat 540
Qy 541 gggaaagagagacattgaggaaccgtgttcaagagaagaccgcctgtgttgaggaggtt 600
Db 541 gggaaagagagacattgaggaaccgtgttcaagagaagaccgcctgtgttgaggaggtt 600
Qy 601 gagaaaccaaaggtgtgggtgaacctactcatatcatcactgaccttactgactactatct 660
Db 601 gagaaaccaaaggtgtgggtgaacctactcatatcatcactgaccttactgactactatct 660
Qy 661 tctctactgacattcttgaaacatttcagggggtggccatattcttcaattatgagctctg 720
Db 661 tctctactgacattcttgaaacatttcagggggtggccatattcttcatatgagctctg 720
Qy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
RESULT 2
US-60-226-176-756
: Sequence 756, Application US/60226176
: GENERAL INFORMATION:
: APPLICANT: Ring, Huijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Townley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GX-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226,176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 756
: LENGTH: 743
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/KEY: misc.feature
: OTHER INFORMATION: GB:HUM2C9X02
US-60-226-176-756
Query Match 99.6%; Score 739.8; DB 55; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Oy 361 agagctctcggcagagcttggccatccacatggtccagtgccagtgccagttctctctt 420
Db 361 agagctctcggcagagcttggccatccacatggtccagtgccagtgccagttctctctt 420
Oy 421 cttgctggagctccctcctcctagtttctctctctctctctctctctctctctctctt 480
Db 421 cttgctggagctccctcctcctagtttctctctctctctctctctctctctctctctt 480
Oy 481 tggaaagaaatggaagagatcccggtttctcctcctcctcctcctcctcctcctcctt 540
Db 481 tggaaagaaatggaagagatcccggtttctcctcctcctcctcctcctcctcctcctt 540
Oy 541 ggggaagaggagcattgagagacgtgttcaagagaaagcccgctgctctgtggagagatt 600
Db 541 ggggaagaggagcattgagagacgtgttcaagagaaagcccgctgctctgtggagagatt 600
Oy 601 gaaaaaaacaaaggtggtgacctactcctacatcactgaccttactgacctactct 660
Db 601 gaaaaaaacaaaggtggtgacctactcctacatcactgaccttactgacctactct 660
Oy 661 tctctactgacattcttgaaacatttcagggtggccatctcttccattatgagttgg 720
Db 661 tctctactgacattcttgaaacatttcagggtggccatctcttccattatgagttgg 720
Oy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
```

```
RESULT 3
US-60-233-468-756
; Sequence 756, Application US/60233468
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; APPLICANT: Valdes, Ana
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-2 P
; CURRENT APPLICATION NUMBER: US/60/233,468
; CURRENT FILING DATE: 2000-09-18
; NUMBER OF SEQ ID NOS: 2488
; SOFTWARE: PERL Program
; SEQ ID NO 756
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:HUM2C9X02
US-60-233-468-756
```

```
Query Match 99.6%; Score 739.8; DB 56; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Oy 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Db 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Oy 121 acttctcagctcctcaaggtctatggcctgtgttccactctgttatttggcctgaacc 180
Db 121 acttctcagctcctcaaggtctatggcctgtgttccactctgttatttggcctgaacc 180
Oy 181 catagtgtgtcgtcagatgaagcagtgaaagcctgaatgattgattcgtgagagaga 240
Db 181 catagtgtgtcgtcagatgaagcagtgaaagcctgaatgattgattcgtgagagaga 240
```

```
Oy 241 gtttctggaagagcattttccacatggctgaaagagataacagagatttggtaggtg 300
Db 241 gtttctggaagagcattttccacatggctgaaagagataacagagatttggtaggtg 300
Oy 301 tgcattgctctgtttcagcatctgtcttgggagtgaggagatggaaaacagagacttac 360
Db 301 tgcattgctctgtttcagcatctgtcttgggagtgaggagatggaaaacagagacttac 360
Oy 361 agagctcctcgggcagagcttggccatccacatggctgcccagtgccagttcctctt 420
Db 361 agagctcctcgggcagagcttggccatccacatggctgcccagtgccagttcctctt 420
Oy 421 cttgctggagctcctcctcctagtttctctctctctctctctctctctctctctt 480
Db 421 cttgctggagctcctcctcctagtttctctctctctctctctctctctctctctt 480
Oy 481 tggaaagaaatggaagagatcccggtttctcctcctcctcctcctcctcctcctt 540
Db 481 tggaaagaaatggaagagatcccggtttctcctcctcctcctcctcctcctcctt 540
Oy 541 ggggaagaggagcattgagagacgtgttcaagagaaagcccgctgctctgtggagagtt 600
Db 541 ggggaagaggagcattgagagacgtgttcaagagaaagcccgctgctctgtggagagtt 600
Oy 601 gaaaaaaacaaaggtggtgacctactcctacatcactgaccttactgacctactct 660
Db 601 gaaaaaaacaaaggtggtgacctactcctacatcactgaccttactgacctactct 660
Oy 661 tctctactgacattcttgaaacatttcagggtggccatctcttccattatgagttgg 720
Db 661 tctctactgacattcttgaaacatttcagggtggccatctcttccattatgagttgg 720
Oy 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
```

```
RESULT 4
US-60-313-371-756
; Sequence 756, Application US/60313371
; GENERAL INFORMATION:
; APPLICANT: Ring, Huijun Z.
; APPLICANT: Malsen, Gareth
; APPLICANT: Townley, David
; APPLICANT: Morris, MacDonald
; TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
; FILE REFERENCE: GX-0013-5 P
; CURRENT APPLICATION NUMBER: US/60/313,371
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 2447
; SOFTWARE: PERL Program
; SEQ ID NO 756
; LENGTH: 743
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: GB:HUM2C9X02
US-60-313-371-756
```

```
Query Match 99.6%; Score 739.8; DB 64; Length 743;
Best Local Similarity 99.7%; Pred. No. 3.3e-209;
Matches 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Oy 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Db 1 tcagaaatattgaagcctgtggtgataaaagcacaatacaatacaatacaataatca 60
Oy 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
Db 61 tgcataatcaggcttagcaaatggaacaaatagtaactctgttctgttctctgtct 120
```

```
QY 121 acttctcagctcctcaaaaggtctatggccctgtgttcaactctgtatttttggcctgaacc 180
|||||
Db 121 acttctcagctcctcaaaaggtctatggccctgtgttcaactctgtatttttggcctgaacc 180
|||||
QY 181 catagtgtgtcgtcgtatgaagcagtgaaagaaagcctcctgattgtatttggaggaga 240
|||||
Db 181 catagtgtgtcgtcgtatgaagcagtgaaagaaagcctcctgattgtatttggaggaga 240
|||||
QY 241 gttttctgaaagagcattttcccaactggctgaaagagctaaacagagagatttggtagtg 300
|||||
Db 241 gttttctgaaagagcattttcccaactggctgaaagagctaaacagagagatttggtagtg 300
|||||
QY 301 tgcattgctgtttcagcatctgtcttgggagtgaggagtgaggagagctaac 360
|||||
Db 301 tgcattgctgtttcagcatctgtcttgggagtgaggagtgaggagagctaac 360
|||||
QY 361 agagctcctcggcagagcttggcccatccacatggctgcccagtgctcagcttctcttt 420
|||||
Db 361 agagctcctcggcagagcttggcccatccacatggctgcccagtgctcagcttctcttt 420
|||||
QY 421 cttgctgggattcctcctcctagtttcttctcttcttcttcttcttcttcttcttctt 480
|||||
Db 421 cttgctgggattcctcctcctagtttcttctcttcttcttcttcttcttcttcttctt 480
|||||
QY 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcctcct 540
|||||
Db 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcctcct 540
|||||
QY 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgcttggaggagtt 600
|||||
Db 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgcttggaggagtt 600
|||||
QY 601 gagaaaccaggggtgggtgacctactcctcctcctcctcctcctcctcctcctcctcctcct 660
|||||
Db 601 gagaaaccaggggtgggtgacctactcctcctcctcctcctcctcctcctcctcctcctcct 660
|||||
QY 661 tcttactgacattcttggaaacatttcagggtggcccatatcttcttcttcttcttcttctt 720
|||||
Db 661 tcttactgacattcttggaaacatttcagggtggcccatatcttcttcttcttcttcttctt 720
|||||
QY 721 ttgttagctcatgtgaagcgggg 743
Db 721 ttgttagctcatgtgaagcgggg 743
|||||
```

RESULT 5

```
US-60-212-657-15
: Sequence 15, Application US/60212657
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000672
: CURRENT APPLICATION NUMBER: US/60/212,657
: CURRENT FILING DATE: 2000-06-19
: NUMBER OF SEQ ID NOS: 303
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 15
: LENGTH: 7609
: TYPE: DNA
: ORGANISM: HUMAN
US-60-212-657-15
```

```
Query Match 98.1%; Score 728.8; DB 54; Length 7609;
Best Local Similarity 99.6%; Pred. No. 2e-205;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
QY 1 tcgaaataattgaagcctgtgtggtgaataaaagcacaatacaatgaataatca 50
|||||
Db 3936 tcgaaataattgaagcctgtgtggtgaataaaagcacaatacaatgaataatca 3995
|||||
```

```
QY 61 tgctaaatcaggcttagcaaatggacaaaatagtaacttcgtttgctgttatctctgtct 120
|||||
Db 3996 tgctaaatcaggcttagcaaatggacaaaatagtaacttcgtttgctgttatctctgtct 4055
|||||
QY 121 acttctcagctcctcaaaagctctatggcctgtgttcaactctgtatttttggcctgaacc 180
|||||
Db 4056 acttctcagctcctcaaaagctctatggcctgtgttcaactctgtatttttggcctgaacc 4115
|||||
QY 181 catagtgtgtcgtcgtatgaagcagtgaaagagccctcctgattgtcttggagagaga 240
|||||
Db 4116 catagtgtgtcgtcgtatgaagcagtgaaagagccctcctgattgtcttggagagaga 4175
|||||
QY 241 gttttctgaaagagcattttcccaactggctgaaagagctaaacagagagatttggtagtg 300
|||||
Db 4176 gttttctgaaagagcattttcccaactggctgaaagagctaaacagagagatttggtagtg 4235
|||||
QY 301 tgcattgctgtttcagcatctgtcttgggagtgaggagtgaggagagctaac 360
|||||
Db 4236 tgcattgctgtttcagcatctgtcttgggagtgaggagtgaggagagctaac 4295
|||||
QY 361 agagctcctcggcagagcttggcccatccacatggctgcccagtgctcagcttctcttt 420
|||||
Db 4296 agagctcctcggcagagcttggcccatccacatggctgcccagtgctcagcttctcttt 4355
|||||
QY 421 cttgctgggattcctcctcctcctagtttcttcttcttcttcttcttcttcttcttcttctt 480
|||||
Db 4356 cttgctgggattcctcctcctcctagtttcttcttcttcttcttcttcttcttcttcttctt 4415
|||||
QY 481 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcctcctcct 540
|||||
Db 4416 tggaaagaaatggaagagatcccgcttctcctcctcctcctcctcctcctcctcctcctcct 4475
|||||
QY 541 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgcttggaggagtt 600
|||||
Db 4476 ggggaagaggagcattgaggaccgtgttcaagagaaagcccgctgcttggaggagtt 4535
|||||
QY 601 gagaaaccaggggtgggtgacctactcctcctcctcctcctcctcctcctcctcctcctcct 660
|||||
Db 4536 gagaaaccaggggtgggtgacctactcctcctcctcctcctcctcctcctcctcctcctcct 4595
|||||
QY 661 tcttactgacattcttggaaacatttcagggtggcccatatcttcttcttcttcttcttctt 719
|||||
Db 4596 tcttactgacattcttggaaacatttcagggtggcccatatcttcttcttcttcttcttctt 4655
|||||
QY 720 gttgttagctcatgtgaagcgggg 743
Db 4656 gttgttagctcatgtgaagcgggg 4679
|||||
```

RESULT 6

```
US-60-213-181-56
: Sequence 56, Application US/60213181
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN PHASE I
: TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000690
: CURRENT APPLICATION NUMBER: US/60/213,181
: CURRENT FILING DATE: 2000-06-22
: NUMBER OF SEQ ID NOS: 261
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 56
: LENGTH: 32768
: TYPE: DNA
: ORGANISM: Human
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (1)...(32768)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-181-56
```

Query Match 98.1%; Score 728.8; DB 54; Length 32768;
Best Local Similarity 99.6%; Pred. No. 4e-205;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 60
DB 29095 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 29154
QY 61 tgctaaatcaggcttagcaaatgacaaaatagtaactctgttgcgtgtatctctgtct 120
DB 29155 tgctaaatcaggcttagcaaatgacaaaatagtaactctgttgcgtgtatctctgtct 29214
QY 121 acttctcctagctctcaaaagctctatggccctgtgttcaactctgtatatttgccctgaaacc 180
DB 29215 acttctcctagctctcaaaagctctatggccctgtgttcaactctgtatatttgccctgaaacc 29274
QY 181 catagtgtgtcgtcatgatgatgaagcagtgaaagagccctgattgatcttggagaga 240
DB 29275 catagtgtgtcgtcatgatgatgaagcagtgaaagagccctgattgatcttggagaga 29334
QY 241 gtttctggaagagcattttcccaactggctgaaagagcctaacaagagatttgtaggtg 300
DB 29335 gtttctggaagagcattttcccaactggctgaaagagcctaacaagagatttgtaggtg 29394
QY 301 tgcattgtcctgtttcagcatctgtcttggggatggggagatgaaagagcagagacttac 360
DB 29395 tgcattgtcctgtttcagcatctgtcttggggatggggagatgaaagagcagagacttac 29454
QY 361 agagctctcggcagagcatttcccaactggctgaaagagcctaacaagagatttgtaggtg 420
DB 29455 agagctctcggcagagcatttcccaactggctgaaagagcctaacaagagatttgtaggtg 29514
QY 421 ctgctcgggattccctcctcctagtttgcgttcttccctcctcctcctcctcctcctt 480
DB 29515 ctgctcgggattccctcctcctagtttgcgttcttccctcctcctcctcctcctcctt 29574
QY 481 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctcctcctcct 540
DB 29575 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctcctcctcct 29634
QY 541 ggggaagagagcattgagagcgtgttcaagagagcccgctgcttctgagagatt 600
DB 29635 ggggaagagagcattgagagcgtgttcaagagagcccgctgcttctgagagatt 29694
QY 601 gagaaacacaaagggtgggtgacccctactcctacatcactgaccttactgactatct 560
DB 29695 gagaaacacaaagggtgggtgacccctactcctacatcactgaccttactgactatct 29754
QY 661 tctcactgacattcttggaaacatttccaggggtgcccatattcttcatatgaagt-ctg 719
DB 29755 tctcactgacattcttggaaacatttccaggggtgcccatattcttcatatgaagt-ctg 29814
QY 720 gttgttagctcatgtgaagcgggg 743
DB 29815 gttgttagctcatgtgaagcgggg 29838

RESULT 7
US-60-248-498-24/c
; Sequence 24, Application US/60248498
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000910
; CURRENT APPLICATION NUMBER: US/60/248,498
; CURRENT FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 264
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 51955

TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(51955)
OTHER INFORMATION: n = A,T,C or G
US-60-248-498-24
Query Match 95.3%; Score 708; DB 57; Length 51955;
Best Local Similarity 97.8%; Pred. No. 8e-199;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1;
QY 1 tcagaaatattgaagcctgtgtgctgaataaaagcacaatacaatacaatgaaatataca 60
DB 46910 TCAGAAATATTGAAGCCTGTGTGGCTGAATAAAGCATACAATAAATGAAATATCA 46851
QY 61 tgctaaatcaggcttagcaaatgacaaaatagtaactctgttgcgtgtatctctgtct 120
DB 46950 TGCTAAATCAGGCTTAGCAAAATGGACAAATAGTAATACTCTGCTTGTATCTCTGCT 46791
QY 121 acttctcctagctctcaaaagctctatggccctgtgttcaactctgtatatttgccctgaaacc 180
DB 46790 ACTTCTCCTAGCTCTCAAAAGGCTATGGCCCTGTGTTCACCTCTGATTTTGGGCTGAAACC 46731
QY 181 catagtgtgtcgtcatgatgatgaagcagtgaaagagccctgattgatcttggagaga 240
DB 46730 CATAGTGGTGTGCATGGATATGAAGCAGTGAAGAGGCCCTGATTGATCTTGGAGAGA 46671
QY 241 gtttctggaagagcattttcccaactggctgaaagagcctaacaagagatttgtaggtg 300
DB 46670 GTTTCTCGAAGAGGCAATTTCCCACTGGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 46611
QY 301 tgcattgtcctgtttcagcatctgtcttggggatggggagatgaaagagcagagacttac 360
DB 46610 TGCATGTGCTGTTTCAGCATCTGTCTTGGGATGGGAGGATGAAACACAGACTTAC 46551
QY 361 agagctcctcggcagagcatttggccctcctcctcctcctcctcctcctcctcctcct 420
DB 46550 AGAGCTTCTCGGCAAGCTTTGGCCCATCCACATGGCTGCCCAGTGTGAGCTTCTCTTT 46491
QY 421 ctgctcgggattccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 480
DB 46490 CTGCTCGGATCTCCCTCCTCTAGTTTCTGTTCTCTCTGTTAGGAATGTTTTCAGCAA 46431
QY 481 tggaaagaaatggagagatccggcgttctcctcctcctcctcctcctcctcctcctcct 540
DB 46430 TGGAAAGAAATGGAAGGAGATCCGGCGTTTCTCCCTCATGACGCTGCGGAAATTTGGGAT 46371
QY 541 ggggaagagagcattgagagcgtgttcaagagagaccccgctgcttctgagagatt 600
DB 46370 GGGGAAGAGAGCATTGAGGACCGTGTTCAGAGAGAACCCCGCTGCTTGTGGAGAGTT 46311
QY 601 gagaaacacaaagggtgggtgacccctactcctacatcactgaccttactgactatct 660
DB 46310 GAGAAACACCAAGGGTGGGTGAACATATCTCTCTACTGACCTTCTCTGGACTGCTCTCC 46251
QY 661 tctcactgacattcttggaaacatttccaggggtgcccatattcttcatatgaagt-ctg 719
DB 46250 TCTCTACTGACATCTTGGAAACATTTTCAGGGGTGCCGAGATCTTTATTATTGGAGTCTCTG 46191
QY 720 gttgttagctcatgtgaagcgggg 743
DB 46190 GTTGTAGCTCATGTGAAGCAGGG 46167

RESULT 8
US-60-248-498-61/c
; Sequence 61, Application US/60248498
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

Mon Apr 22 08:41:44 2002

```
Oy 661 tcttactgacattcttgaaacatttcagggtgagccatattcttctcattatgaat-ctg 719
Db 89712 TCTCTACTGACATCTCTGGAACATTTAGGGGTGCCAGATCTTTTATTTGGAGTCTTG 89653

Oy 720 gttgttagctcatgtgaagcggg 743
Db 89652 GTTGTAGCTCATGTGAAGCAGG 89629

RESULT 10
US-60-245-227-9/c
: Sequence 9, Application US/60245227
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000876
: CURRENT APPLICATION NUMBER: US/60/245,227
: CURRENT FILING DATE: 2000-11-03
: NUMBER OF SEQ ID NOS: 129
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 225532
: TYPE: DNA
: ORGANISM: HUMAN
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)...(225532)
: OTHER INFORMATION: n = A,T,C or G
US-60-245-227-9

Query Match 95.3%; Score 708; DB 57; Length 225532;
Best Local Similarity 97.8%; Pred. No. 1.6e-198;
Matches 728; Conservative 0; Mismatches 15; Indels 1; Gaps 1:

Oy 1 tcagaaatatttgaaagcctgtgtgctgaataaaagcatacaatacaatacaataca 60
Db 93367 TCAGAAATATTTGAAAGCCTGTGTGCTGAATAAAAGCATACAATACAATAATCA 93308

Oy 61 tctaaatcaggcttagcaaatgacaaaatagtaactctgttctgttatctctgtct 120
Db 93307 TCCTAAATCAGCTTAGCAAAATGACAAAATAGTAACCTTCGTTGCTGTTAICTGTCT 93248

Oy 121 acttcttagctctcaaggcttatggccctgtatggccctgtatgttctgttctgtct 180
Db 93247 ACTTCTTAGCTCTCAAGGCTATGGCCCTGTGTCACTCTGTATTTTGGCCCTGAAACC 93188

Oy 181 catagtgtgtcgtcatgatgaagcagtgtaaggaagccctgattgatcttgagagaga 240
Db 93187 CATAGTGTGTGTCATGATGAAGCAGTGAAGGAGCCCTGATTGATCTTGGAGAGGA 93128

Oy 241 gttttctgaaagagcattttccactggctgaaagagcctaacagagatttgtaggtg 300
Db 93127 GTTTTCTGGAAGAGCATTTTCCCACTGGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 93068

Oy 301 tcatgtcctgtttcagcatctgtcttggtgagatggagagatgaaacacagagacttac 360
Db 93067 TGCATGTGCTGTTTTCAGCATCTGTCTTGGGATGGGAGGATGTAACACAGAGACTTAC 93008

Oy 361 agagctcctcgggagagcttggcccatccacatggctggccagtgtagcttctcttt 420
Db 93007 AGAGCTCTCGGGAGAGCTTGGCCCATCCACATGGCTGCCAGTGTCAGCTTTCCTCTT 92948

Oy 421 ctgtcctgggagatcctcctcagttcttcttctctctgttagaattgttttcagcaa 480
Db 92947 CTTGCCCTGGGATCTCCCTCCTAGTTTCGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 92888

Oy 481 tggaaagaaatggaagagatccgcgtttctccctcatgaagcgtcggaattttggagat 540
Db 92887 TGGAAAGAAATGGAAGAGATCCGCGCTTCTCCCTCATGACCTGCGGAAATTTGGGAT 92828
```

```
Oy 541 ggggaagagagcattgaggaccgtgttcaagagaagcccgctgttctgtggaggagt 600
Db 92827 GGGGAAGAGGAGCATTGAGGACCGTGTTCAGAGGAAGCCGCTGCTTGTGGAGAGTT 92768

Oy 601 gagaataaccgaaggtggtgaccctactccatatacactgacctactgactactatct 660
Db 92767 GAGAAAAACCAAGGGTGGGTGAACATACTCTATCACTGACCTTTCCTGGAGACTGCTCTCC 92708

Oy 661 tctctactgacattcttgaaacatttcagggtgagccatattcttctcattatgaat-ctg 719
Db 92707 TCTCTACTGACATCTCTGGAACATTTTCAGGGGTGCCAGATCTTTTATTTGGAGTCTCTG 92648

Oy 720 gttgttagctcatgtgaagcggg 743
Db 92647 GTTGTAGCTCATGTGAAGCAGG 92624

RESULT 11
US-09-488-127-59
: Sequence 59, Application US/09488127
: GENERAL INFORMATION:
: APPLICANT: Thomann, Hans-Ulrich
: APPLICANT: FitzGerald, Michael
: TITLE OF INVENTION: RAPID DETERMINATION OF GENE STRUCTURE
: TITLE OF INVENTION: USING CDNA SEQUENCE
: FILE REFERENCE: 2709.1005-000
: CURRENT APPLICATION NUMBER: US/09/488,127
: CURRENT FILING DATE: 2000-01-20
: NUMBER OF SEQ ID NOS: 96
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 59
: LENGTH: 8437
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (1)...(8437)
: OTHER INFORMATION: n = A,T,C or G
US-09-488-127-59

Query Match 83.8%; Score 622.6; DB 18; Length 8437;
Best Local Similarity 93.3%; Pred. No. 9.7e-174;
Matches 694; Conservative 0; Mismatches 45; Indels 5; Gaps 4:

Oy 1 tcagaaatatttgaaagcctgtgtgctgaataaaagcatacaatacaatacaataca 60
Db 1391 tcagaaatatttg-agcctgtgtgactgaataaaagcatacaatacaatacaataca 1449

Oy 61 tctaaatcaggcttagcaaatgacaaaatagtaactctgttctgttctgtctgtct 120
Db 1450 atctaagtcaggcttagtaaatggacaaaacagtgacctcatctgtctttaaactgtatct 1509

Oy 121 acttcttagctctcaaggctctatggccctgtgttcaactctgtattttggcctgaaacc 180
Db 1510 cttttctnag-tctcaaaaatctatggccctgtgttcaactctgtattttggcctgaaacg 1568

Oy 181 catagtgtgtcgtcatgatgaagcagtgaaagagcctgattgatcttgagagaga 240
Db 1569 catgtgtgtcgtcatgatgaatgaagtgtggaagagcctgattgatcttgagagaga 1628

Oy 241 gttttctggaagagcattttccacatggctgaaagagcctaacagagatttgtaggtg 300
Db 1629 gttttctggaagagcattttccacatggctgaaagagcctaacagagatttgtaggtg 1688

Oy 301 tgcattgtcctgtttcagcatctgtcttggtgagatgggagagatggaaaaacagagacttac 360
Db 1689 tgcagtgctgtttcagcatctgtcttggtgagatgggagagatggaaaaac--agactcagc 1746

Oy 361 agagctcctcgggagagatcccgctttctccctcatgaagcgtgtcagcttctctcttt 420
Db 1747 agagctcctcgggagagatcccgctttctccctcatgaagcgtgtcagcttctctcttt 1806
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 09:32:08 ; Search time 4168.4 Seconds
(without alignments)
1278.330 Million cell updates/sec

Title: US-09-763-292-3

Perfect score: 323

Sequence: 1 cccctgaattgtcaacaa.....aaattcatagatcatctttt 323

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_hlg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pl:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sv:**

13: gb_un:**

14: gb_vl:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_om:**

20: em_or:**

21: em_ov:**

22: em_pat:**

23: em_ph:**

24: em_pi:**

25: em_ro:**

26: em_sts:**

27: em_sv:**

28: em_un:**

29: em_vl:**

30: em_hgo_hum:**

31: em_hgo_inv:**

32: em_hgo_rod:**

33: em_hlg_hum:**

34: em_hlg_inv:**

35: em_hlg_rod:**

36: em_hlg_other:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Query Match	Length	DB	ID	Description
1	323	100.0	323	9	HUM2C9X05	L16881 Homo sapien
2	310.4	96.1	143087	2	AL359672	AL359672 Homo sapi
3	275.2	85.2	173154	2	AL133513	AL133513 Homo sapi
4	275.2	85.2	205791	2	AL583836	AL583836 Homo sapi
5	263	81.4	348	9	HUM2C18X06	L16874 Homo sapien
6	263	81.4	156492	2	AL157835	AL157835 Homo sapi
7	222.4	68.9	143087	2	AL359672	AL359672 Homo sapi
8	208.8	64.6	201340	2	AC018872	AC018872 Homo sapi
9	208.8	64.6	210821	2	AC013318	AC013318 Homo sapi
10	194	60.1	595	4	RABP4501	M74204 Rabbit cyto
11	192	59.4	1854	6	AR071576	AR071576 Sequence
12	192	59.4	1854	9	HUM2C9X05	M61855 Human cytoc
13	190.4	58.9	1473	22	E10631	E10631 Human cdna
14	190.4	58.9	1473	22	E10853	E10853 cdna encodi
15	190.4	58.9	1814	9	S45963	S45963 putative cy
16	190.4	58.9	1826	9	HUMCYPH2	D00173 Homo sapien
17	190.4	58.9	1843	6	E02279	E02279 Human liver
18	190.4	58.9	1845	9	HUM2C9X05	M61857 Human cytoc
19	190.4	58.9	1852	6	AR071579	AR071579 Sequence
20	188.8	58.5	1441	9	HUMCYPMP8	M21940 Human cytoc
21	188.8	58.5	1576	9	HUMCYPMP	M15331 Human liver
22	188.8	58.5	1577	9	HUMCYPMPA	M21939 Human cytoc
23	180.6	55.9	1892	6	AR071581	AR071581 Sequence
24	179.8	55.7	147838	2	AC083888	AC083888 Rattus no
25	179.2	55.5	1276	9	HUMCYP2C17	M61858 Human cytoc
26	177.6	55.0	1346	9	HSCP450	X65962 H.sapiens m
27	177.6	55.0	1746	6	AR071575	AR071575 Sequence
28	177.6	55.0	1746	9	HUMCYP2C19	M61854 Human cytoc
29	176.6	54.7	330	22	E11555	E11555 DNA probe t
30	176	54.5	1444	6	E14930	E14930 Artificial
31	176	54.5	1473	22	E10866	E10866 cdna encodi
32	176	54.5	1669	6	E14931	E14931 Human mrna
33	176	54.5	2395	9	HUMSV450A	L07093 Human cytoc
34	176	54.5	168457	2	AC078913	AC078913 Mus muscu
35	168	52.0	1473	22	E10865	E10865 cdna encodi
36	168	52.0	1995	9	HUMCYP2C18	M61856 Human cytoc
37	168	52.0	2009	6	AR071577	AR071577 Sequence
38	168	52.0	2258	6	AR071580	AR071580 Sequence
39	168	52.0	2258	9	HUM2C18	M61853 Human cytoc
40	166	51.4	218	10	RATCY45M6	M18361 Rat cytochr
41	164.8	51.0	1307	4	SSSSCC103	Z93098 S.scrofa cy
42	164.8	51.0	1758	4	SSC15D09	Z93100 S.scrofa cy
43	161.2	49.9	294	10	RATP450PB1	M24237 Rat cytochr
44	158.6	49.1	1751	10	MUSCP450	D17674 Mouse mrna
45	156.8	48.5	1856	10	RATCYPM1	J02657 Rat cytochr

ALIGNMENTS

RESULT

1

HUM2C9X05

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

HUM2C9X05 323 bp DNA PRI 08-FEB-1999
Homo sapiens cytochrome P450C9 (CYP2C9) gene, exon 7.

L16881
CYP2C9; cytochrome P450; mephenytoin 4-hydroxylase.
5 of 7

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

human

REFERENCE
AUTHORS
Goldstein, J.A., Raucy, J.L., Blaisdell, J.A., Faletto, M.B. and Romkes, M.

TITLE
JOURNAL
MEDLINE
REFERENCE
Cloning and expression of complementary DNAs for multiple members of the human cytochrome P4501C subfamily
Biochemistry 30, 3247-3255 (1991)
91182740
2 (bases 1 to 323)

AUTHORS de Morais, S.M., Schweikl, H., Blaisdell, J. and Goldstein, J.A.
 TITLE Gene structure and upstream regulatory regions of human CYP2C9 and CYP2C18
 JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
 MEDLINE 93326116
 FEATURES
 source Location/Qualifiers
 1..323
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 /tissue_lib="EMBL3 library from J.A. Goldstein"
 70..257
 /gene="CYP2C9"
 /note="coding sequence identical to Lue 359 variant (clone 25)"
 /citation=[1]
 /number=7
 /evidence-experimental
 BASE COUNT 83 a 88 c 59 g 93 t
 ORIGIN
 Query Match 100.0%; Score 323; DB 9; Length 323;
 Best Local Similarity 100.0%; Pred. No. 1.5e-84;
 Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 cccctgaattgctacacaaatgcccattttctcttttccatcagtttttacttg 60
 Db 1 CCCCTGAATTCCTACACAAATGTGCCATTTCTCTTTTCCATCAGTTTACTTGTG 60
 Oy 61 tcttatcagctaaagtcacagaagattgaacgtgtgattggcagaaaccggagccct 120
 Db 61 TCTTATCAGCTAAAGTCACGAAGAGATTGAAGCTGTGATTGGCAGAAACCGAGCCCT 120
 Oy 121 gcattgcaagacagagccacatgccctacacagatgctgtggtgcacagatccagagat 180
 Db 121 GCATGCAAGACAGGAGCCACATGCCCTACACAGATGCTGTGTTGGCAGAGTCCAGAGT 180
 Oy 181 accttgaccttctcccaccagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
 Db 181 ACCTTGACCTTCTCCCCACAGCGCTGCCCATGCAGTGCAGCTGTGATTAATTCAGAA 240
 Oy 241 actatctcattcccaagtgattgtttctcctacacagcagcagcagcagcagcagcagc 300
 Db 241 ACTATCTCATCCCAAGTAAAGTTGTTTCTCTACACTGCAACTCCATGTTTTCGAAGT 300
 Oy 301 cccaaattcatagatcattttt 323
 Db 301 CCCAAATTCATAGTATCATTTT 323
 RESULT 2
 LOCUS AL359672
 DEFINITION Homo sapiens chromosome 10 clone RP11-208C17, 27-JUN-2001
 PROGRESS ***, 2 unordered pieces.
 ACCESSION AL359672
 VERSION AL359672.17 GI:14575223
 KEYWORDS HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 143087)
 Johnson, C.
 Direct Submission
 Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 Requests: clonerquest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14529836.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC

Web site: <http://www.sanger.ac.uk>
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA208C17
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 142666 bases at least Q40
 Consensus quality: 142843 bases at least Q30
 Consensus quality: 142920 bases at least Q20
 Insert size: 142987; sum-of-contigs
 Insert size: 156404; agarose-fp
 Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
 coverage: 7.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 47005: contig of 47005 bp in length
 * 47006 47105: gap of 100 bp
 * 47106 143087: contig of 95982 bp in length.

FEATURES

source

1..143087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="10"
 /clone="RP11-208C17"
 /clone_lib="RPC1-11.1"
 /note="assembly_fragment:00494
 clone_end:SP6
 vector_side:left"
 47106..143087
 /note="assembly_fragment:00967"
 26155 c 27788 g 47797 t 100 others

misc_feature

misc_feature

BASE COUNT 41247 a 26155 c 27788 g 47797 t 100 others
 ORIGIN

Query Match 96.1%; Score 310.4; DB 2; Length 143087;
 Best Local Similarity 99.4%; Pred. No. 1.4e-80;
 Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1 cccctgaattgctacacaaatgcccattttctcttttccatcagtttttacttg 60
 Db 99506 CCCCTGAATTCCTACACAAATGTGCCATTTCTCTTTTCCATCAGTTTACTTGTG 99565
 Oy 61 tcttatcagctaaagtcacagaagattgaacgtgtgattggcagaaaccggagccct 120
 Db 99566 TCTTATCAGCTAAAGTCCAGGAAGAGATTGAAGCTGTGATTGGCAGAAACCGAGCCCT 99625
 Oy 121 gcattgcaagacagagccacatgccctacacagatgctgtggtgcacagatccagagat 180
 Db 99626 GCATGCAAGACAGGAGCCACATGCCCTACACAGATGCTGTGTTGGCAGAGTCCAGAGAT 99685
 Oy 181 accttgaccttctcccaccagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
 Db 99686 ACATGTACCTTCTCCCCACAGCGCTGCCCATGCAGTGCAGTGCAGTGCAGTGCAGT 99745
 Oy 241 actatctcattcccaagtgattgtttctcctacacagcagcagcagcagcagcagcagc 300
 Db 99746 ACTATCTCATCCCAAGTAAAGTTTGTTCCTCTACACTGCAACTCCATGTTTTCGAAGT 99805
 Oy 301 cccaaattcatagatcattttt 323
 Db 99806 CCCAAATTCATAGTATCATTTT 99829

```

RESULT 3
AL133513
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-400G3, HTG 04-JUL-2001
PROGRESS ***, in ordered pieces.
ACCESSION AL133513
VERSION AL133513.11 GI:14575067
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Brown.J.
Direct Submission
Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14456168.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA400G3
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 172989 bases at least Q40
Consensus quality: 173136 bases at least Q30
Consensus quality: 173145 bases at least Q20
Insert size: 173154; sum-of-contigs
Insert size: 116005; 33.1% error; agarose-fp
Quality coverage: 6.48x in Q20 bases; sum-of-contigs Quality
coverage: 9.80x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..173154
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-400G3"
/clone_lib="RPCI-11.2"
/misc_feature 1..173154
/note="assembly_fragment:00748
clone_end:SP6
vector_side:right"
BASE COUNT 49163 a 32272 c 33707 g 58012 t
ORIGIN
Query Match 85.2%; Score 275.2; DB 2: Length 173154;
Best Local Similarity 92.6%; Pred. No. 3.4e-70;
Matches 300; Conservative 0; Mismatches 23; Indels 1; Gaps 1:
QY 1 cccctgaattgtacacaaatgacatttctctcttccatcagtttttactgtg 60
DB 51799 CCCCTGAATTGCTAGACAAATGTTCCATTCTCTCTTTTCCATCAGTTCTTACTGTG 51858
QY 61 tcttatcagtaagtccagggaagattgaacgtgtgattggcagaaccggagccct 120
DB 51859 TCTTGTTCAGTAAAGTCCAGGAAGATTGAACGTGTCGTGGCAGAAACCGAGCCCT 51918
QY 121 gcatcgaagacagagaccacatccctacacagatctgtgtgacagaggtccgaagt 180
DB 51919 GCATGCAGGACAGGGGCCACATCCCTACACAGATGCTGTGGTGCACGAGGTCCAGAT 51978

```

```

QY 181 accctgaccttctcccccaccagcctgcccccatcagtcagctgacattaaattcagaa 240
DB 51979 ACATGCACCTATCCCCACCACGAGCTGCCCATGAGTGTGACGTAAATTCAGAA 52038
QY 241 actatctcattcccaaggtaagttgttctctctcacgtcgaactcattgtttcgaagt 300
DB 52039 ACTACCTCATTCGCAAGGTAAGTTGTCTCTCTACACTGCAACTCCATGTTCTTTAT 52098
QY 301 cc-caaatcatagatcatctttt 323
DB 52099 CTTCAATTCACAGATGATGATCTT 52122
-----
RESULT 4
AL583836
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-466J14, HTG 13-JUN-2001
PROGRESS ***, 2 unordered pieces.
ACCESSION AL583836
VERSION AL583836.11 GI:14455940
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Mashreghi-Mohammadi.M.
Direct Submission
Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:14148873.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA466J14
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 205236 bases at least Q40
Consensus quality: 205448 bases at least Q30
Consensus quality: 205560 bases at least Q20
Insert size: 205691; sum-of-contigs
Insert size: 188405; 9.1% error; agarose-fp
Quality coverage: 11.31x in Q20 bases; sum-of-contigs Quality
coverage: 12.57x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 88240: contig of 88240 bp in length
* 88241 88340: gap of 100 bp
* 88341 205791: contig of 117451 bp in length.
FEATURES
source
1..205791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-466J14"
/clone_lib="RPCI-11.2"
/misc_feature 1..88240
/note="assembly_fragment:00191
fragment_chain:1"
misc_feature 88341..205791

```

/note="assembly_fragment:01703
fragment_chain:1
clone_end:T7
vector_side:right"

BASE COUNT 56694 a 38262 c 40890 g 69845 t 100 others
ORIGIN

Query Match 85.2%; Score 275.2; DB 2; Length 205791;
Best Local Similarity 92.6%; Pred. No. 3.4e-70;
Matches 300; Conservative 0; Mismatches 23; Indels 1; Gaps 1;
QY 1 cccctgaattgctacaacaaatggtccatttttctctttccatcagtttttacttggtg 50
Db 164775 CCCCTGAATTGCTAGAACAAATGTCATTTCCTCTTTTCCATCAGTCTTACTTIG 164834
QY 61 tcttatcagctaaagtccagaagaagattgaacgtgtgattgagcagaacccgagccct 120
Db 164835 TCTTGTGACGCTAAAGTCAGGAAGAGATTGAACGTGTGTCGTTGGCAGAAACGGAGCCCT 164894
QY 121 gcatgcaagacagagagcccatgcccacacagatgctgtggtgacagaggtccagagat 180
Db 164895 GCATGCAGGACAGGCGGCACATGCCCTACACAGATGCTGTGGTGCACAGAGTCCAGAGAT 164954
QY 181 acctgacctctcccacacagcctgcccacatgcagtgacctgtgacattaaattcagaa 240
Db 164955 ACATCGACCTCATCCACACGAGCTGCCATGCAGTGACCTGTGACGTTAAATTCAGAA 165014
QY 241 actatctcattcccaagtgattgtttctctacactgacactcaactcactcatttccgaagt 300
Db 165015 ACTACCTCATCTCCAGGTAAAGTTGTTCTCTCTACACTGCAACTCCTCATGTTCTTTAT 165074
QY 301 cc-caaattcattagatcattttt 323
Db 165075 CCTCAAAATTCACAGATGATGTTCT 165098

RESULT 5
LOCUS HUM2C18X06
DEFINITION Homo sapiens cytochrome P4502C18 (CYP2C18) gene, exon 7.
ACCESSION L16874
VERSION L16874.1 GI:291604
KEYWORDS CYP2C18; cytochrome P450; mephenytoin 4-hydroxylase.
SEGMENT 6 of 8
SOURCE Homo sapiens (library: EMBL3 library from J.A. Goldstein) liver
DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Goldstein,J.A., Raucy,J.L., Blaisdell,J.A., Faletto,M.B. and Romkes,M.
TITLE Cloning and expression of complementary DNAs for multiple members of the human cytochrome P4501C subfamily
JOURNAL Biochemistry 30, 3247-3255 (1991)
MEDLINE 91182740
REFERENCE 2 (bases 1 to 348)
AUTHORS de Morais,S.M., Schweikl,H., Blaisdell,J. and Goldstein,J.A.
TITLE Gene structure and upstream regulatory regions of human CYP2C9 and CYP2C18
JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
MEDLINE 93326116
REFERENCE 3 (sites)
AUTHORS Romkes,M., Faletto,M.B., Blaisdell,J.A., Raucy,J.L. and Goldstein,J.A.
TITLE Correction: Cloning and expression of complementary cDNAs for multiple members of the human cytochrome P4501C subfamily
JOURNAL Biochemistry 32, 1390-1390 (1993)
MEDLINE 93192243
FEATURES
Location/Qualifiers
1..348,
/organism="Homo sapiens"

/db_xref="taxon:9606"
/tissue_type="liver"
/tissue_lib="EMBL3 library from J.A. Goldstein"
order(L16873,1,206..325,<1..100)
/gene="CYP2C18"
/number=6
101..288
/gene="CYP2C18"
/citation={1}
/citation={3}
/number=7
/evidence-experimental
BASE COUNT 89 a 97 c 62 g 100 t
ORIGIN

Query Match 81.4%; Score 263; DB 9; Length 348;
Best Local Similarity 91.7%; Pred. No. 7.2e-67;
Matches 289; Conservative 0; Mismatches 25; Indels 1; Gaps 1;
QY 3 cctgaattgctacaacaaatggtccatttttctctttccatcagtttttacttggtgc 62
Db 34 CCTGAATTGCTACACAAATGTCATTTCCTCTTTTCCATCATTTCTTACTTGTGC 93
QY 63 ttatcagctaaagtccagaagaagattgaacgtgtgattgagcagaacccgagccctgc 122
Db 94 TTATCAGCTAAAGTCAGGAAGAGATTGAATGTGTAGTTGGCAGAAACCGGAGCCCTGT 153
QY 123 atgcaagacagagagccacatgcccacacagatgctgtggtgacagaggtccagagatac 182
Db 154 ATGCAGGACAGGAGTACATGCCCTACACAGATGCTGTGGTGCACAGATCCAGAGATAC 213
QY 183 ctgacctctcccacacagcctgcccacatgcagtgacctgtgacattaaattcagaaac 242
Db 214 ATGACCTCTCTCCACCAACCTGCCATGCAGTGACCTGTGATGTTAAATTCAAAAC 273
QY 243 tatctcattcccaagtgattgtttctctacactgacactcaactcatttccgaagt-c 301
Db 274 TACCTCATCTCCCAAGGTAAAGTTGTTCTCTCTACACTGATCTCCATCTCTCAAGTCC 333
QY 302 ccaaattcattagat 316
Db 334 CCAAATTCATAGTAT 348

RESULT 6
LOCUS AL157835 156492 bp DNA
DEFINITION Homo sapiens chromosome 10 clone RP11-361K9, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION AL157835
VERSION AL157835.9 GI:14626943
KEYWORDS HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 156492)
AUTHORS Blakey,S.
TITLE Direct Submission
JOURNAL Submitted (05-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK
E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 2001 this sequence version replaced gi:14586042.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA361K9
----- Summary Statistics
Assembly program: XGAP4; version 4.5

```

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 156104 bases at least Q40
Consensus quality: 156214 bases at least Q30
Consensus quality: 156256 bases at least Q20
Insert size: 156492; sum-of-contigs
Insert size: 150975; 13.7% error; agarose-fp
Quality coverage: 8.34x in Q20 bases; sum-of-contigs Quality
coverage: 9.69x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES             Location/Qualifiers
     source            1..156492
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="10"
                        /clone="RP11-361K9"
                        /clone_lib="RP11-11.2"
     misc_feature      1..156492
                        /note="assembly_fragment:03685
                        clone_end:SP6
                        vector_side:left"
BASE COUNT           43071 a 30007 c 31050 g 52364 t
ORIGIN
Query Match          81.4%; Score 263; DB 2; Length 156492;
Best Local Similarity 91.7%; Pred. No. 1.3e-66;
Matches 289; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

Oy 3 cctgaattgctacacaaatgtgccattttctctttccatcagttttacttgctc 62
|||||
Db 140291 CCTGAATTGCTACAAATGTGCCATTTCTCTCTTTCATCAITTCITACITGTGTC 140350

Oy 63 ttatcagctaaagtccaggaagatgtgaactgtgattggcagaaaccgagccctgc 122
|||||
Db 140351 TTATCAGCTTAAAGTCCAGGAAGAGATTGAATGTGTAGTTGGCAGAAACCGGAGCCCTGT 140410

Oy 123 atgcaagacagagccacatgcctcacagatgctgtggtgacagaggtccagagatac 182
|||||
Db 140411 ATGCAGGACAGAGTCACTACATGCCCTTACACAGATGCTGTGGTGCAGGATCCAGAGATAC 140470

Oy 183 ctgtgacctctcccaccagcctgccccatgcagtgacctgtgacattaaattcagaaac 242
|||||
Db 14047; ATTGACCTCTCCCCACCACCACTGCCCATGCAGTGACCTGTGATGTTAAATTCAAAAAC 140530

Oy 243 tatctcattcccgaagttaagtgtttctctctacactgcaactccatgttttcgaagt-c 301
|||||
Db 140531 TACCTCATCCCCAAGTAAAGTGTGTTCTCTCTACACTACATCTCCATGCTCTTCAAGTCC 140590

Oy 302 ccaaatccatagtat 316
|||||
Db 140591 CCAAATTCATAGTAT 140605

RESULT              7
LOCUS               AL359672/c
DEFINITION          Homo sapiens chromosome 10 clone RP11-208C17, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
ACCESSION           AL359672
VERSION             AL359672.17 GI:14575223
KEYWORDS            HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE              human.
ORGANISM            Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE           1 (bases 1 to 143087)
AUTHORS             Johnson,C.
TITLE               Direct Submission

```

JOURNAL

Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14529836.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA208C17
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 142666 bases at least Q40
Consensus quality: 142843 bases at least Q30
Consensus quality: 142920 bases at least Q20
Insert size: 142987; sum-of-contigs
Insert size: 156404; agarose-fp
Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
coverage: 7.82x in Q20 bases; agarose-fp

COMMENT

* NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
be preserved.
* 1 47005: contig of 47005 bp in length
* 47006 47105: gap of 100 bp
* 47106 143087: contig of 95982 bp in length.
Location/Qualifiers
1..143087
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-208C17"
/clone_lib="RP11-11.1"
1..47005
/note="assembly_fragment:00494
clone_end:SP6
vector_side:left"
47106..143087
/note="assembly_fragment:00967"
BASE COUNT 41247 a 26155 c 27788 g 47797 t 100 others
ORIGIN

Query Match 68.9%; Score 222.4; DB 2; Length 143087;
Best Local Similarity 83.2%; Pred. No. 1.2e-54;
Matches 253; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

FEATURES

source

misc_feature

misc_feature

BASE COUNT 41247 a 26155 c 27788 g 47797 t 100 others
ORIGIN

Oy 2 cctgaattgctacacaaatgtgccattttctctttccatcagttttacttgctg 61
|||||
Db 18256 CCATGAATTGCTATGACAAATGTTCATATATCTTCGTTCCATCAGTCTTCTTGTGT 18197
Oy 62 cttatcagctaaagtccaggaagatgtgaactgtgattggcagaaaccgagccctg 121
|||||
Db 18196 CTGTGACGCTAAAGTCCAGGAAGAGATTGATCATGTAATTGGCAGACACAGGAGCCCTG 18137
Oy 122 catgcaagacagagccacatgcctcacagatcgtgtggtgacagaggtccagagata 181
|||||
Db 18136 CATGCAGGATAGGAGCCACATGCCTTACACTGATGCTGTAGTGCACGAGATCCAGAGATA 18077
Oy 182 ccttgacctctcccaccagcctgccccatgcagtgacctgtgacattaaattcagaaa 241
|||||
Db 18076 CAGTGACCTTGTGCCCCCGGTGTCGCCCATGCAGTGACCACTGATACTAAGTTTCAGAAA 18017
Oy 242 ctatctcattcccgaagttaagtgtttctctctacactgcaactccatgttttcgaagtc 301
|||||

```

Db 18016 CTACCTCATCCCAAGTAAGCTGTTCTCTTACACTATATTTCTGTACTTCTGAAATT 17957
QY 302 ccaa 305
Db 17956 TCCA 17953

RESULT 8
AC018872
LOCUS
DEFINITION Homo sapiens chromosome 2 clone RP11-140M22, WORKING DRAFT
ACCESSION AC018872
VERSION
KEYWORDS
SOURCE HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 201340)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 201340)
Waterston,R.H.
Direct Submission
Submitted (21-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 18, 2001 this sequence version replaced gi:8954228.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0140M22
----- Summary Statistics -----
Sequencing vector: M13; 71%
Chemistry: Dye-primer ET; 71% of reads
Chemistry: Dye-terminator Big Dye; 29% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 187240 bases at least Q40
Consensus quality: 190955 bases at least Q30
Consensus quality: 193344 bases at least Q20
Insert size: 178000; agarose-fp
Insert size: 198640; sum-of-contigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 4.13 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1402: contig of 1402 bp in length
* 1403 1502: gap of unknown length
* 1503 2841: contig of 1339 bp in length
* 2842 2941: gap of unknown length
* 2942 4465: contig of 1524 bp in length
* 4466 4565: gap of unknown length
* 4566 7341: contig of 2776 bp in length
* 7342 7441: gap of unknown length
* 7442 9248: contig of 1807 bp in length
* 9249 9349: gap of unknown length
* 9349 11809: contig of 2461 bp in length
* 11810 15022: contig of 3113 bp in length
* 15023 15122: gap of unknown length

```

```

* 15123 18880: contig of 3758 bp in length
* 18881 18980: gap of unknown length
* 18981 21664: contig of 2684 bp in length
* 21665 25032: contig of 3268 bp in length
* 25033 30071: contig of 4939 bp in length
* 30072 35416: contig of 5245 bp in length
* 35417 35517: gap of unknown length
* 35518 40752: contig of 5236 bp in length
* 40753 40852: gap of unknown length
* 40853 45905: contig of 5053 bp in length
* 45906 46005: gap of unknown length
* 46006 50784: contig of 4779 bp in length
* 50785 56904: gap of unknown length
* 56905 57004: contig of 6020 bp in length
* 57005 64526: contig of 7522 bp in length
* 64527 72095: gap of unknown length
* 72096 72195: contig of 7469 bp in length
* 72196 79372: gap of unknown length
* 79373 79472: contig of 7177 bp in length
* 79473 87932: gap of unknown length
* 87933 88033: contig of 8460 bp in length
* 88034 97114: gap of unknown length
* 97115 109392: contig of 9082 bp in length
* 109393 109492: gap of unknown length
* 109493 119348: contig of 12178 bp in length
* 119349 119447: gap of unknown length
* 119448 133975: contig of 9855 bp in length
* 133976 134075: gap of unknown length
* 134076 150216: contig of 14528 bp in length
* 150217 150316: contig of 16141 bp in length
* 150317 164328: gap of unknown length
* 164329 164429: contig of 14012 bp in length
* 164430 178475: gap of unknown length
* 178476 178575: contig of 14046 bp in length
* 178576 201340: contig of 22766 bp in length.

FEATURES
Location/Qualifiers
source
1..201340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone="RP11-140M22"
misc_feature
1..1402
/note="assembly_name:Contig11"
misc_feature
1503..2841
/note="assembly_name:Contig12"
misc_feature
2942..4465
/note="assembly_name:Contig13"
misc_feature
4566..7341
/note="assembly_name:Contig15"
misc_feature
7442..9248
/note="assembly_name:Contig16"
misc_feature
9349..11809
/note="assembly_name:Contig17"
misc_feature
11910..15022
/note="assembly_name:Contig18"
misc_feature
15123..18980
/note="assembly_name:Contig19"
misc_feature
18981..21664
/note="assembly_name:Contig21"
vector_side:left
21765..25032
/note="assembly_name:Contig20"
misc_feature
25133..30071
/note="assembly_name:Contig22"
misc_feature
30172..35416
/note="assembly_name:Contig23"

```

```

misc_feature 35517..40752
/notes="assembly_name:Contig24"
misc_feature 40853..45905
/notes="assembly_name:Contig25"
misc_feature 46006..50784
/notes="assembly_name:Contig26"
misc_feature 50885..56904
/notes="assembly_name:Contig27"
misc_feature 57005..64526
/notes="assembly_name:Contig28"
misc_feature 64627..72095
/notes="assembly_name:Contig29"
misc_feature 72196..79372
/notes="assembly_name:Contig30"
misc_feature 79473..87932
/notes="assembly_name:Contig31"
misc_feature 88033..97114
/notes="assembly_name:Contig32"
misc_feature 97215..109392
/notes="assembly_name:Contig33"
misc_feature 109493..119347
/notes="assembly_name:Contig34"
misc_feature 119448..133975
/notes="assembly_name:Contig35"
misc_feature 134076..150216
/notes="assembly_name:Contig36"
misc_feature 150317..164328
/notes="assembly_name:Contig37"
misc_feature 164429..178474
/notes="assembly_name:Contig38"
misc_feature 178575..201340
/notes="assembly_name:Contig39"

BASE COUNT 61711 a 36248 c 35995 g 64637 t 2749 others
ORIGIN

Query Match 64.6% Score 208.8; DB 2: Length 201340;
Best Local Similarity 82.5%; Pred. No. 1.3e-50;
Matches 264; Conservative 0; Mismatches 52; Indels 4; Gaps 2;

QY 3 cctgaattgtacaaataatgtccattttctctttccatcagttttactgtgtc 62
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116688 CCTGAATGCTACGACAAATGTGCCATTTTCTCCTTTTTCATCAGTCTCCTGTGTC 116747

QY 63 ttatcagctaaagtccagggaagattgaactgtgtattgacagaaaccgagccctgc 122
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116748 TTGTCATCTAAAGTCCAGGAGAGACGGACCATGTCAGTTGGCAGACTGGAGCCCTGC 116807

QY 123 atgcaagacggagccacatgccctacacagatgctgtgtgacaggtccagagatac 182
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116808 ATGCAGGACAGGAGCCACATGCCCTACACAGAAGCCATGGTGCATGAGTCCAGAGACAC 116867

QY 183 cttagaccttctccaccagcctgcccacatgcagtcagtcacattaaattcagaaac 242
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116868 ---TGACCTACCCACCAATGTGCCCATGTGACCTGACCTGACATTAATTTAGAAAC 116924

QY 243 tatctattcccaaggttaattgtttctctacacatgcacactccatttttcgaagt-c 301
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116925 TACCTCTCCCAAGGTAAGCTTTTCTCTAGACTGTGGCTGTGCTTTGATGTC 116984

QY 302 ccaattatcatgatacttt 321
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 116985 CCAAAATTCACAGTAATGTTT 117004

RESULT 9
AC013318/c AC013318 210821 bp DNA HTG 23-SEP-2000
LOCUS Homo sapiens chromosome 2 clone RP11-358N5 map 2, WORKING DRAFT
DEFINITION SEQUENCE, 22 unordered pieces.
ACCESSION AC013318
VERSION AC013318.5 GI:10280844
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

```

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 210821)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-358N5
2 (bases 1 to 210821)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 23, 2000 this sequence version replaced gi:8072541.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3773
Center clone name: 358_N_5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 196258 bases at least Q40
Consensus quality: 203137 bases at least Q30
Consensus quality: 206079 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 208721; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 5808: contig of 5808 bp in length
* 5809 5908: gap of 100 bp
* 5909 7108: contig of 1200 bp in length
* 7109 7208: gap of 100 bp
* 7209 8378: contig of 1170 bp in length
* 8379 8478: gap of 100 bp
* 8479 9873: contig of 1395 bp in length
* 9874 9973: gap of 100 bp
* 9974 12585: contig of 2612 bp in length
* 12586 12685: gap of 100 bp
* 12686 16506: contig of 3821 bp in length
* 16507 16606: gap of 100 bp
* 16607 20660: contig of 4054 bp in length
* 20661 20760: gap of 100 bp
* 20761 24298: contig of 3538 bp in length

* 24299 24398: gap of 100 bp
* 24399 29876: contig of 5478 bp in length
* 29877 29976: gap of 100 bp
* 29977 36607: contig of 6631 bp in length
* 36608 36707: gap of 100 bp
* 36708 43142: contig of 6435 bp in length
* 43143 43242: gap of 100 bp
* 43243 49210: contig of 5968 bp in length
* 49211 49310: gap of 100 bp
* 49311 55880: contig of 6570 bp in length
* 55881 55980: gap of 100 bp
* 55981 63850: contig of 7870 bp in length
* 63851 63950: gap of 100 bp
* 63951 71005: contig of 7055 bp in length
* 71006 71105: gap of 100 bp
* 71106 80701: contig of 9596 bp in length
* 80702 80801: gap of 100 bp
* 80802 126993: contig of 46192 bp in length
* 126994 127093: gap of 100 bp
* 127094 139959: contig of 12866 bp in length
* 139960 140059: gap of 100 bp
* 140060 155880: contig of 15821 bp in length
* 155881 155980: gap of 100 bp
* 155981 170719: contig of 14739 bp in length
* 170720 170819: gap of 100 bp
* 170820 200516: contig of 29697 bp in length
* 200517 200616: gap of 100 bp
* 200617 210821: contig of 10205 bp in length.

FEATURES

source

1. 210821
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="2"
/clone="RP11-358N5"
/clone_lib="RP11 Human Male BAC"
1. 5808
/note="assembly_fragment
clone_end:SP6
vector_side:left"

/note="assembly_fragment"
7209. 8378
/note="assembly_fragment"
8479. 9873
/note="assembly_fragment"
9974. 12585
/note="assembly_fragment"
12686. 16506
/note="assembly_fragment"
16607. 20660
/note="assembly_fragment"
20761. 24298
/note="assembly_fragment"
24399. 29876
/note="assembly_fragment"
29977. 36607
/note="assembly_fragment"
36708. 43142
/note="assembly_fragment"
43243. 49210
/note="assembly_fragment"
49311. 55880
/note="assembly_fragment"
55981. 63850
/note="assembly_fragment"
63951. 71005
/note="assembly_fragment"
71106. 80701
/note="assembly_fragment"
80802. 126993
/note="assembly_fragment"
127094. 139959

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

exon

/note="assembly_fragment"
140060. 155880
/note="assembly_fragment"
155981. 170719
/note="assembly_fragment"
170820. 200516
/note="assembly_fragment"
200617. 210821
/note="assembly_fragment
clone_end:T7
vector_side:right"

BASE COUNT 67328 a 37637 c 36968 g 66777 t 2111 others
ORIGIN

Query Match 64.6%; Score 208.8; DB 2; Length 210821;
Best Local Similarity 82.5%; Pred. No. 1.3e-50;
Matches 264; Conservative 0; Mismatches 52; Indels 4; Gaps 2;

Qy 3 cctgaattgtacacaaatgcccattttctctcttccatcagtttttactgtgtc 62
Db 25158 CCTGAAGTGTACGACAAATGTCCATTTTCTCTTTTCATCAGTCTCACTTGTGTC 25099
Qy 63 ttatcagctaaagtccaggaagatgaaagtgtgattgacagaaacggagccctgc 122
Db 25098 TTGTCATCTAAAGTCCAGGAGAGACGACATGTCAGTTGGCAGACACTGGAGCCCTGC 25039
Qy 123 atgcagaagaggagccatgcctcacagatgctgtgtgcacaggtccagagatac 182
Db 25038 ATGCAGCAGGAGGACCATGCCCTACAGAGCCATGTCATGATGAGGTCAGAGACAC 24979
Qy 183 ctgacacctctccaccagcctgcacgtgcagtgacacgtgacatttaataaccagaaac 242
Db 24978 ---TGACCTCACCCCAATGTGCCCATGTGCTGACCTCTGACATTAATTTAGAAAC 24922
Qy 243 tatctcattcccaaggttaagtgtttctctctacactgcaactccatgttttcgaagt-c 301
Db 24921 TACCTCTCCCAAGGTAAAGTTGTTCTCTCTAGACTGTGGCTCTATGCTCTTGATGTC 24862
Qy 302 ccaaatctcatgattcattt 321
Db 24861 CCAATTCACAGTATTGTTT 24842

RESULT 10

RABP4501

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SEGMENT

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

FEATURES

source

Location/Qualifiers

1. 595

/organism="Oryctolagus cuniculus"

/strain="New Zealand White"

/db_xref="taxon:9986"

/cell_type="hepatocyte"

/tissue_type="liver"

/dev_stage="adult"

/germline

162. 349

```
/gene="CYP2C5"
/number=7
/evidence=experimental
BASE COUNT 160 a 130 c 110 g 195 t
ORIGIN

Query Match 60.1%; Score 194; DB 4; Length 595;
Best Local Similarity 81.0%; Pred. No. 1.7e-46;
Matches 238; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

Oy 2 cccatgaattgtacacaaaatgtgccattttctctcttttccatcagtttttactttgtg 60
||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 93 CCTAAATGTATGACAGCTGTGCCAATTTCTCATTTCCCGTTATTTCTTACTTAIG 152

Oy 61 tcttatcagctaaagtccagaagaattgaacgtgtgattggcagaaaccggagccct 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 153 TCTTATCAGCTAGGTCCTCAGGAGGATTTGAGCTGTGATTCGACAGACCGAGCCCCI 212

Oy 121 gcatgaagacaggagccacatgccctacacagatgctgtgtgcacgaggtccagat 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 213 GCATGCAGGACAGGAGCGCATGCCCTACACGGATGCTGTGATACATGAGATCCAGAGAT 272

Oy 181 accctgaacttctcccacacagcctgcccctgacgtgacctgtgacattaaattcagaa 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 273 TCATCGACCTCTTCCCACTAACCTGCCCATGCCATGACCTGACCTGATTTAGATTTCAGAA 332

Oy 241 actatctcattcccagaagtgtgtttctcctacactgaactccatatttt 294
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 333 ACTACTTTATCCCTAAGGTATCTTGTTCTTTTACATGACCTCAGTCTCT 386

RESULT 11
AR071576 1854 bp DNA PAT 18-FEB-2000
LOCUS AR071576 Sequence 4 from patent US 5912120.
DEFINITION AR071576
ACCESSION AR071576
VERSION AR071576.1 GI:7222464
KEYWORDS Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1854)
AUTHORS Goldstein,J.A. and De Morais,S.M.F.
TITLE Cloning, expression and diagnosis of human cytochrome P450 2C19:
the principal determinant of s-mephenytoin metabolism
JOURNAL Patent: US 5912120-A 4 15-JUN-1999;
FEATURES
source Location/Qualifiers
1..1854
/organism="unknown"
BASE COUNT 513 a 424 c 381 g 536 t
ORIGIN

Query Match 59.4%; Score 192; DB 6; Length 1854;
Best Local Similarity 100.0%; Pred. No. 7.3e-46;
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaaccggagccctgcagtc 126
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 971 CAGCTAAAGTCCAGGAGAGATTGAACGTGTGNTGGCAGAAACCGAGCCCTGCATGC 1030

Oy 127 aagacagagccacatgccctacacagatgctgtgtgacagagtcacagatacccttg 186
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1031 AAGACAGGAGCCACATATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACCTTG 1090

Oy 187 acctctccccaccagcctgccccatgcagctgcagctgtgacattaaattcagaaactatc 246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1091 ACCTTCTCCCCACAGCCTGCCCCATGCAGTGCACCTGTGACATTAAATTCAGAAACTATC 1150

Oy 247 tcattcccaagg 258
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db 1151 TCATTCCCAAGG 1162

RESULT 13
E10631 standard; RNA: HUM; 1473 BP.
ID E10631
XX AC E10631;
XX SV E10631.1
XX
DT 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE Human cDNA encoding cytochrome P4502C9.
XX
KW JP 1996027196-A/5.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
```

```

RN [1]
RP 1-1473
RA Funae Y., Imaoka S., Matsuki Y., Hayashi K., Yabusaki Y.;
RT *ANTIBODY RECOGNIZING CYTOCHROME P450C9 ORIGINATED FROM MAN*;
RL Patent number JP1996027196-A/5, 30-JAN-1996.
RL SUMITOMO CHEM CO LTD.
XX
CC OS Homo sapiens (human)
CC PN JP 1996027196-A/5
CC PD 30-JAN-1996
CC PF 13-JUL-1994 JP 1994161551
CC PI FUNAE YOSHIHIKO, IMAOKA SUSUMU, MATSUKI YASUSHI, HAYASHI KOJI,
CC YABUSAKI YOSHIYASU
CC PC C07K16/18.C12N15/09.G01N33/53.G01N33/53//C12N1/19.C12N9/02,
CC (C12N1/19,
CC C12R1:865);
CC CC strandedness: Double;
CC CC topology: Linear;
CC CC Key Location/Qualifiers
CC FH Key
CC FT source 1..1473
CC FT /organism="Homo sapiens"
CC FT /tissue_type="liver"
CC FT 1..1473
CC FT /product="cytochrome P450C9"
XX
FH Key Location/Qualifiers
FH FT source 1..1473
FH FT /db_xref="taxon:9606"
FH FT /organism="Homo sapiens"
XX
SQ Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other;

Query Match 58.9%; Score 190.4; DB 22; Length 1473;
Best Local Similarity 99.5%; Pred. No. 2.1e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaacccgagccctgcatgc 126
DB 959 CAGCTAAAGTCCAGGAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCAATGC 1018

QY 127 aagacaggagccacatgccctacacagatgctgtgtgcacaggtccagagataccttg 186
DB 1019 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1078

QY 187 accttctcccacagcctgccctacagtcagtgacctgtgacattaaattcagaaactatc 246
DB 1079 ACCTTCTCCCACAGCCTGCCCTACAGTGCACGTGACCTGTGACATTAAATTTCAGAAACTATC 1138

QY 247 tcattcccaagg 258
DB 1139 TCATTCCCAAGG 1150

RESULT 14
E10E53
ID E10853 standard; RNA; HUM; 1473 BP.
XX
AC E10853;
XX
SV E10853.1
XX
DT 07-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE cDNA encoding human cytochrome P450.
XX
KW JP 1996056695-A/2.
XX
OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN [1]
RP 1-1473
RA Hayashi K., Sakaki T., Yabusaki Y., Komai K., Kaneko H., Nakatsuka I.;
RT *METHOD FOR EVALUATING SAFETY*;
RL Patent number JP1996056695-A/2, 05-MAR-1996.
RL SUMITOMO CHEM CO LTD.
XX
CC OS Homo sapiens (human)
CC PN JP 1996056695-A/2
CC PD 05-MAR-1996
CC PF 15-JUL-1994 JP 1994164184
CC PR 20-JUL-1993 JP 93P 201120, 30-JUL-1993 JP 93P 208279,
CC 17-JUN-1994 JP 94P 136053
CC PI HAYASHI KOJI, SAKAKI TOSHIYUKI, YABUSAKI YOSHIYASU,
CC KOMAI KOICHIRO,
CC KANEKO HIDEO, NAKATSUKA IWAO
CC PC C12Q1/02.C12M1/34.C12Q1/26;
CC CC strandedness: Double;
CC CC topology: Linear;
CC CC hypothetical: No;
CC CC anti-sense: No;
CC FH Key Location/Qualifiers
CC FH source 1..1473
CC FH /organism="Homo sapiens"
CC FT CDS 1..1473
CC FT /product="human cytochrome P450 2C9"
XX
FH Key Location/Qualifiers
FH FT source 1..1473
FH FT /db_xref="taxon:9606"
FH FT /organism="Homo sapiens"
XX
SQ Sequence 1473 BP; 413 A; 345 C; 319 G; 396 T; 0 other;

Query Match 58.9%; Score 190.4; DB 22; Length 1473;
Best Local Similarity 99.5%; Pred. No. 2.1e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 67 cagctaaagtcaggagagattgaacgtgtgattggcagaaacccgagccctgcatgc 126
DB 959 CAGCTAAAGTCCAGGAGAGATTGAACGTGTGATTGGCAGAAACCGGAGCCCTGCAATGC 1018

QY 127 aagacaggagccacatgccctacacagatgctgtgtgcacaggtccagagataccttg 186
DB 1019 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGCACGAGGTCCAGAGATACATTG 1078

QY 187 accttctcccacagcctgccctacagtcagtgacctgtgacattaaattcagaaactatc 246
DB 1079 ACCTTCTCCCACAGCCTGCCCTACAGTGCACGTGACCTGTGACATTAAATTTCAGAAACTATC 1138

QY 247 tcattcccaagg 258
DB 1139 TCATTCCCAAGG 1150

RESULT 15
S46963
LOCUS S46963 1814 bp mRNA PRI 05-AUG-1999
DEFINITION putative CYP2C9 gene [human, liver, mRNA, 1814 nt].
ACCESSION S46963
VERSION S46963.1 GI:258514
KEYWORDS human liver.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1814)
AUTHORS Ohguya,S., Komori,M., Ohi,H., Shiramatsu,K., Shinriki,N. and

```

TITLE Kamatani, T.
Six-base deletion occurring in messages of human cytochrome P-450
in the CYP2C subfamily results in reduction of tolbutamide
hydroxylase activity
JOURNAL Biochem. Int. 27 (6), 1073-1081 (1992)
MEDLINE 93075249
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsg 117179] from the original journal article.
This sequence comes from Fig. 1.
COMMENT Author also gives sequence for another cytochrome P-450 that
contains a 6-base deletion of nucleotides 783-788.
FEATURES
 Source
 1..1814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 gene
 1..1814
 CDS
 3..1436
 /gene="putative CYP2C9 gene"
 /note="This sequence comes from Fig. 1"
 /codon_start=1
 /product="cytochrome P-450"
 /protein_id="AAB23864.2"
 /db_xref="GI:5705937"
 /translation="MLLSLWRQSGRGKLPPTPLPVLVIGNLIQIKDISKLTNL
 SKVYGPVTLFGLKPIVVLHGVEAVKEALIDGEEFSGRGIFPLAERANRGFGIVFS
 NGKVKWEIRFSLMTLRNFGMKRSIEDRVOEAEARCLVEELRKTASPCDPIFILGCA
 PCNVICSIIFHKRFDYKDOQLNLMEKLNENIKILSSPMQICNNFSPIDYFPCTHN
 KLLKNVAFMKSYILEKVKHESMDMNPQDFIDCFMKMEKEKHNPQSEFTIESLEN
 TAVDLFGAGTETTTSTLYALLLLKHPETAKVOEETERVIGHNRSPCMODRSHMPY
 TDVAVHEVQRTIDILPTSLPHAVICDIFRNYLIPKGTITLISLTVLHDNKEFPNPE
 MFDPHHFLOEGNFKSKSYFMPFSGAKRICVGEALAGMELFLFSLTILQNFNLKSLVD
 PKNLDTTPVNVGFAFVPPFYQLCFIPV"
BASE COUNT 517 a 413 c 369 g 515 t
ORIGIN

Query Match 58.9%; Score 190.4; DB 9; Length 1814;
Best Local Similarity 99.5%; Pred. No. 2.2e-45;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 67 cagctaaagtccaggaagattgaacgtgtgattggcagaacccgagagccctgcacatgc 126
 |||||
Db 922 CAGCTAAAGTCAGGAAGAGATTGAACTGTGTGATGGCAGAAACCGGACCCCTGCATGC 981
 |||||
QY 127 aagacaggagccacatgcccctacacagatgctgtgtgacagaggtccagagataccttg 186
 |||||
Db 982 AAGACAGGAGCCACATGCCCTACACAGATGCTGTGTGTCACAGAGTCCAGAGATACATG 1041
 |||||
QY 187 acctctccccaccagcctgccctcagtcagtgacattgacattaaattcagaaactatc 246
 |||||
Db 1042 ACCTTCTCCACAGCGCTGCCCATGACCTGTGACATTAATTCAGAAACTATC 1101
 |||||
QY 247 tcattcccaagg 258
 |||||
Db 1102 TCATTCCCAAGG 1113

Search completed: April 19, 2002, 09:34:57
Job time: 7997 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:10:39 : Search time 130.44 Seconds
(without alignments)
560.813 Million cell updates/sec

Title: us-09-763-292-3
Perfect score: 333
Sequence: 1 cccctgaattgctaacacaa.....aaattcatgatacttttt 333

Scoring table: IDENTITY.MSC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0

Maximum Match 1000

Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/lna/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/lna/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/lna/5A_COMB.seq.*
4: /cgn2_6/ptodata/2/lna/5B_COMB.seq.*
5: /cgn2_6/ptodata/2/lna/PTGUS_COMB.seq.*
6: /cgn2_6/ptodata/2/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	59.4	1854	1	US-08-201-118-4
2	192	59.4	1854	2	US-08-201-118-4
3	192	59.4	1854	3	US-08-201-118-4
4	190.4	58.9	1852	1	US-08-201-118-10
5	190.4	58.9	1852	2	US-08-201-118-10
6	190.4	58.9	1852	3	US-08-201-118-10
7	188.8	58.5	1851	2	US-08-194-981E-4
8	188.8	58.5	1851	3	US-08-194-981E-4
9	180.6	55.9	1892	1	US-08-201-118-14
10	180.6	55.9	1892	2	US-08-201-118-14
11	180.6	55.9	1892	3	US-08-201-118-14
12	177.6	55.0	1746	1	US-08-201-118-2
13	177.6	55.0	1746	2	US-08-201-118-2
14	177.6	55.0	1746	3	US-08-201-118-2
15	168	52.0	2009	1	US-08-201-118-6
16	168	52.0	2009	2	US-08-201-118-6
17	168	52.0	2009	3	US-08-201-118-6
18	168	52.0	2258	1	US-08-201-118-12
19	168	52.0	2258	2	US-08-201-118-12
20	168	52.0	2258	3	US-08-201-118-12
21	153.6	47.6	1829	1	US-08-201-118-8
22	153.6	47.6	1829	2	US-08-201-118-8
23	153.6	47.6	1829	3	US-08-201-118-8
24	100.4	31.1	1482	2	US-08-194-981E-1
25	100.4	31.1	1482	3	US-08-194-981E-1
26	91	28.2	8779	2	US-08-750-703-1
27	83.4	25.8	1737	2	US-08-750-703-2

Sequence 5, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 17, Appli
Sequence 14, Appli
Sequence 21, Appli
Sequence 23, Appli
Sequence 16, Appli
Sequence 19, Appli
Sequence 13, Appli
Sequence 20, Appli
Sequence 21, Appli
Sequence 18, Appli
Sequence 15, Appli
Sequence 24, Appli
Patent No. 5508199
Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-201-118-4
: Sequence 4, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: INVENTOR: JOYCE A. MARJORIE
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Hourie and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA: US/08201118
: APPLICATION NUMBER: 05/08201118
: FILING DATE: 09-APR-1992
: CLASSIFICATION: 433
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 329-0400
: TELEFAX: (415) 329-0424
: INFORMATION FOR SEQ ID NO. 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1854 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-201-118-4

Query Match 59.4%, Score 192; DB 1; Length 1854;
Best Local Similarity 100.0%; Pred. No. 2 (1e-5);
Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OTHER INFORMATION: /note- *Corresponds to positions -12 to -1
OTHER INFORMATION: for 25 of Figure 2.*
S-08-238-821B-4

3
 PCT-US95-05744-4
 : Sequence 4, Application PC/TUS9505744
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : APPLICANT: DE MORAIS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CITROCHROME P450 2C19: THE PRINCIPAL DETERMINANT
 : TITLE OF INVENTION: OF S-MEPHINTOIN METABOLISM
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Townsend and Townsend Khourie and Crew
 : STREET: 379 Lytton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : CURRENT APPLICATION IN RELEASE \$1.0, Version \$1.25
 : FILING DATE: 09-APR-1992
 : APPLICATION NUMBER: PCT/US95/05744
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/238,821
 : FILING DATE: 06-MAY-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 22-FEB-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Dow, Karen B.
 : REGISTRATION NUMBER: 29,684
 : REFERENCE/DOCKET NUMBER: 15280-192-1-1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 326-2400
 : TELEFAX: (415) 326-2422
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1854 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: cDNA

.; MOLECULE TYPE: CDNA
PCT-US95-05744-10

Query Match	58.9%	Score 190.4	DB 5	Length 1852	
Best Local Similarity	99.5%	Pred. No. 6.8e-51			
Matches 191	Conservative	0	Mismatches 1	Indels 0	Gaps 0
67	cacgtcaaatgccgaagaagattgacgctgtgatttcgcgaataccgagccctcgatgc	126			
68					
69	cacgtcaaatgccgaagaagattgacgctgtgatttcgcgaataccgagccctcgatgc	1028			
127	aagcagcagagccacatgccctaccacagatctgtgtgcacgaggttcacagatacctgt	186			
129	aagcagcagagccacatgccctaccacagatctgtgtgtgcacgaggttcacagatacctgt	1088			
187	acctctctcccccacgagctgcgcccatcgagtcgacatctaattccgaagaatc	246			
1089	acctctctcccccacgagctgcgcccatcgagtcgacatctaattccgaagaatc	1148			
247	tcattcccaagg	258			
1149	tcattcccaagg	1150			

```

1  RESULT
2  7
3  US-08-194-981E-4
4  : Sequence 4, Application US/08194981E
5  : Patent No. 5886157
6  :
7  : GENERAL INFORMATION:
8  :
9  : APPLICANT: GUENGERICH, F. Peter
10 : APPLICANT: GUO, Zuyu
11 : APPLICANT: GUO, Zuyu
12 : APPLICANT: GILLAM, Elizabeth M. J.
13 : TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
14 : TITLE OF INVENTION: HUMAN
15 : TITLE OF INVENTION: CYTOCHROME P450
16 : NUMBER OF SEQUENCES: 68
17 :
18 : CORRESPONDENCE ADDRESS:
19 : ADDRESSEE: NEEDLE & ROSENBERG, P.C.
20 : STREET: Suite 1200, 127 Peachtree Street, NE
21 : CITY: Atlanta
22 : STATE: Georgia
23 : COUNTRY: USA
24 :
25 : ZIP: 30303-1811
26 :
27 : COMPUTER READABLE FORM:
28 : MEDIUM TYPE: Floppy disk
29 : COMPUTER: IBM PC compatible
30 : OPERATING SYSTEM: PC-DOS/MS-DOS
31 : CURRENT APPLICATION DATE: 11/01/98, Version 41.25
32 :
33 : APPLICATION NUMBER: US/08/194,981E
34 : FILING DATE: February 10, 1994
35 : CLASSIFICATION: 435
36 : ATTORNEY/AGENT INFORMATION:
37 : NAME: Elizabeth Selby
38 : REGISTRATION NUMBER: 38,298
39 : REFERENCE/DOCKET NUMBER: 22000.0022
40 : TELEPHONE: (404) 688-0770
41 : TELEPHONE: (404) 688-0770
42 : TELEFAX: (404) 688-9880
43 :
44 : INFORMATION FOR SEQ ID NO: 4:
45 : SEQUENCE CHARACTERISTICS:
46 : LENGTH: 1419 base pairs
47 : TYPE: nucleic acid
48 : STRANDEDNESS: single
49 : TOPOLOGY: linear
50 : MOLECULE TYPE: DNA (genomic)
51 : ANTI-SENSE: NO
52 :
53 : US-08-194-981E-4

```

Query Match 58.5%; Score 188.8; DB 2; Length 1419;
Best Local Similarity 99.0%; Pred. No. 2e-50;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaagtcagagagattggaacgtgtagtgagagaaacccgagccctcgcacg 126
Db 959 CAGCTAAAGTCAGGAGAGATTGGAACGTGTGATGGCAGAAACCGGAGCCCTGCATGC 1018
Oy 127 agagcagagagcctacccctacacagatgctgtgtagcagagagcagagataccttg 186
Db 1019 AAGACAGAGGCGCACATGCCCTACACAGATGCTGTGTGCGAGGCTCCAGAGATGCATTG 1078
Oy 187 accttctccccacagcctgccccacagcagcagcctgacatttaattcagaactatc 246
Db 1079 ACCTTCTCCCCACAGCCTGCCCATCGAGCTGCGACCTGACCTGTGACATTAATTAATTCAGAAATATC 1138
Oy 247 tcattcccaagg 258
Db 1139 TCATTCCCAAGG 1150

RESULT 9
US-08-201-118-14
: Sequence 14, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: NUMBER OF INVENTORS: 14
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201,118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Selby
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: YES
US-08-201-118-14

Query Match 55.9%; Score 180.6; DB 1; Length 1892;
Best Local Similarity 95.3%; Pred. No. 8.7e-48;
Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 67 cagctaaagtcagagagattggaacgtgtagtgagagaaacccgagccctcgcacg 126
Db 1001 CAGCTAAAGTCAGGAGAGATTGGAACGTGTGATGGCAGAAACCGGAGCCCTGCATGC 1060

Query Match 58.5%; Score 188.8; DB 2; Length 1419;
Best Local Similarity 99.0%; Pred. No. 2e-50;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaagtcagagagattggaacgtgtagtgagagaaacccgagccctcgcacg 126
Db 905 CAGCTAAAGTCAGGAGAGATTGGAACGTGTGATGGCAGAAACCGGAGCCCTGCATGC 964
Oy 127 agagcagagagcctacccctacacagatgctgtgtagcagagagcagagataccttg 186
Db 965 AAGACAGAGGCGCACATGCCCTACACAGATGCTGTGTGCGAGGCTCCAGAGATGCATTG 1024
Oy 187 accttctccccacagcctgccccacagcagcagcctgacatttaattcagaactatc 246
Db 1025 ACCTTCTCCCCACAGCCTGCCCATCGAGCTGCGACCTGACCTGTGACATTAATTAATTCAGAAATATC 1084
Oy 247 tcattcccaagg 258
Db 1085 TCATTCCCAAGG 1096

RESULT 8
US-08-194-981E-3
: Sequence 3, Application US/08194981E
: Patent No. 5688157
: GENERAL INFORMATION:
: APPLICANT: GRENKE, F. Peter
: APPLICANT: GRENKE, F. Peter
: APPLICANT: SANDHU, Punam
: APPLICANT: GILLAM, Elizabeth M. J.
: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF
: TITLE OF INVENTION: HUMAN
: TITLE OF INVENTION: CYTOCHROME P450
: NUMBER OF SEQUENCES: 68
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: Suite 1200, 127 Peachtree Street, NE
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303-1811
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/194,981E
: FILING DATE: February 10, 1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Elizabeth Selby
: REGISTRATION NUMBER: 38,298
: REFERENCE/DOCKET NUMBER: 22000.0022
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-8880
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1591 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEICAL: NO
: AMTI-SEQSE: NO
US-08-194-981E-3

Query Match 58.5%; Score 188.8; DB 2; Length 1591;
Best Local Similarity 99.0%; Pred. No. 2.1e-50;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

07 127 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagatacttg 186
 Db 1061 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGACAGAGGTCCAGAGATACATTG 1120
 07 187 accttctccacacagcctccacacagcctccacacagcctccacacagcctccacac 245
 Db 1121 ACTCTCTCCACACAGCCTCCACACAGCCTCCACACAGCCTCCACACAGCCTCCACAC 1180
 07 247 tcatctcccaagg 258
 Db 1181 TCATNCCCAAGG 1192

RESULT 10
 : Sequence 14, Application US/08238821B
 : Patent No. 5912120
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMEK-SPARKS, Marjorie
 : APPLICANT: DE MORALS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
 : OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, 8th Floor
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: US
 : ZIP: 94111
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patent in Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/238,821B
 : FILING DATE: 06-MAY-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 22-FEB-1994
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: 435
 : FILING DATE: 06-MAY-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Liebeschuetz, Joe
 : REGISTRATION NUMBER: 37,505
 : REFERENCE/DOCKET NUMBER: 15280-1921100S
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (650) 326-2400
 : TELEFAX: (650) 326-2422
 : INFORMATION: 14
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1892 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : FEATURE:
 : NAME/KEY: Region
 : LOCATION: 1..41
 : OTHER INFORMATION: /note= *Corresponds to positions -41 to -1
 : OTHER INFORMATION: for 2c of Figure 2.
 US-08-238-821B-14

Query Match 55.9% Score 180.6; DB 2; Length 1892;
 Best Local Similarity 95.3%; Pred. No. 8.7e-48;

Matches 183: Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 07 67 cagctaaagtcacagagagattgaacatgtatgttgacagaaacagagccctgcattgc 126
 Db 1001 CAGCTAAAGTCCAGAGAGAGATTGAACGTGTATTTGCGAAGAACCGGACCCCTGCATGC 1060
 07 127 aagacagagccacatgccctacacagatgctgtgtgacagaggtccagagatacttg 186
 Db 1061 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGACAGAGGTCCAGAGATACATTG 1120
 07 187 accttctccacacagcctccacacagcctccacacagcctccacacagcctccacac 245
 Db 1121 ACTCTCTCCACACAGCCTCCACACAGCCTCCACACAGCCTCCACACAGCCTCCACAC 1180
 07 247 tcatctcccaagg 258
 Db 1181 TCATNCCCAAGG 1192
 RESULT 11
 : PCT-US95-05744-14
 : Sequence 14, Application PC/TUS9505744
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMEK-SPARKS, Marjorie
 : APPLICANT: DE MORALS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
 : OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Townsend and Townsend Khourile and Crew
 : STREET: 379 Fulton Avenue
 : CITY: Palo Alto
 : STATE: California
 : COUNTRY: US
 : ZIP: 94301
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patent in Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US95/05744
 : FILING DATE:
 : CLASSIFICATION:
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/238,821
 : FILING DATE: 06-MAY-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/201,118
 : FILING DATE: 22-FEB-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 07/864,962
 : FILING DATE: 09-APR-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Dow, Karen B.
 : REGISTRATION NUMBER: 29,684
 : REFERENCE/DOCKET NUMBER: 15280-192-1-1
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 326-2400
 : TELEFAX: (415) 326-2422
 : INFORMATION: 14
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1892 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : HYPOTHETICAL: YES
 PCT-US95-05744-14

Query Match 55.9%; Score 180.6; DB 5; Length 1892;
Best Local Similarity 95.3%; Pred. No. 8.7e-48;
Matches 183; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

07 67 cagctaaagtcaggagagattgaacgtgctgattgacgagaaacggagccctgcacgc 126
DB 1001 CAGCTAAAGTCAGGAGAGATTGAAGCTGTGATTGGCAGAAACGGAGCCCTGCATGC 1060

07 127 aagacagagccacatgcctctacacagatgctgtgtggtgacgaggtccagagataccttg 186
DB 1061 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGTCACGAGGTCCAGAGATACATGC 1120

07 187 accctctcccacagctgcccacatgcacgtgacgtgacattcaataatcagaactatc 246
DB 1121 ACCTCTCTCCCACAGCTGCCCATGCACGTCGATGAGTTAAATTCAGAACTACC 1180

07 247 tcatctcccaagg 258
DB 1181 TCATCTCCCAAGG 1192

RESULT 12
US-08-201-118-2
: Sequence 1, Application 05/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourie and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201.118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-192-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1746 base pairs
: TYPE: nucleic acid
: STRANDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Region
: LOCATION: 1..5
: OTHER INFORMATION: /note= "Corresponds to positions 15 to 1
: US-08-201-118-2

DB 964 CAGCTAAAGTCAGGAGAGATTGAACGTGCTGATTGGCAGAAACGGAGCCCTGCATGC 1023
DB 127 aagacagagccacatgcctctacacagatgctgtgtggtgacgaggtccagagataccttg 186
DB 1024 AGACAGAGCCACATGCCCTACACAGATGCTGTGTGTCACGAGGTCCAGAGATACATGC 1083
DB 187 accctctcccacagctgcccacatgcacgtgacgtgacattcaataatcagaactatc 246
DB 1084 ACCTCTCTCCCACAGCTGCCCATGCACGTCGATGAGTTAAATTCAGAACTACC 1143

07 247 tcatctcccaagg 258
DB 1144 TCATCTCCCAAGG 1155

RESULT 13
US-08-238-821B-2
: Sequence 2, Application US/08238821B
: Patent No. 5912120
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
: TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
: NUMBER OF SEQUENCES: 61
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: 400 California Center, 8th Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: US
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/238.821B
: FILING DATE: 06-MAY-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201.118
: FILING DATE: 22-FEB-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 15280-1921100S
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 326-2400
: TELEFAX: (650) 326-2422
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1746 base pairs
: TYPE: nucleic acid
: STRANDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: Region
: LOCATION: 1..5
: OTHER INFORMATION: /note= "Corresponds to positions 15 to 1
: US-08-238-821B-2

```

Query Match      55.04; Score 177.6; DB 5; Length 1746;
Best Local Similarity 95.34; Red. No. 7.3e-47;
Matches 183; Conservative 0; Missmatches 9; Indels 0; Gaps 0;

Oy 67 cagcctaaagtccagaagagatgaactgtgtgattgcagaaacccagagccctgcattgc 126
Db 964 CAGCTAAAGTCCAGAGAGATTGAACGTGTGATTGCGAGAACCCGGAGCCCTGCATCG 1023

Oy 127 agcagcaggagccatccctctacacagatgctgtgtgcacagaggtccagagataccttg 186
Db 1024 AGGACAGAGGGCCACATGCCCTACACAGATGCTGTGTGCAGCGAGTCCAGAGATCAATCG 1083

Oy 187 acccttccccacacgcctgcgcccctgcagtgacctgtgacattataatccagaacattc 246
Db 1084 ACCTCATCCGACACAGCTGTGCCCTCTGCAGTACCTGTAGCTGTTAAATCAGAACTACC 1143

Oy 247 tcattcccaagg 258
Db 1144 TCATCCCAAG 1155

```

```

1  RESULT 15
2  US-08-201-118-6
3  : Sequence 6, Application US/08201118
4  : Patent No. 5786191
5  : GENERAL INFORMATION:
6  : APPLICANT: GOLDSTEIN, Joyce A.
7  : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
8  : TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C18
9  : NUMBER OF INVENTION: SUBFAMILY
10 : NUMBER OF SEQUENCES: 44
11 : CORRESPONDENCE ADDRESS:
12 : ADDRESSEE: Townsend and Townsend Kourile and Crew
13 : STREET: 379 Lytton Avenue
14 : CITY: Palo Alto
15 : STATE: California
16 : COUNTRY: US
17 : ZIP: 94301
18 : COMPUTER READABLE FORM:
19 : MEDIUM TYPE: Floppy disk
20 : OPERATING SYSTEM: IBM PC compatible
21 : SOFTWARE: PatentId Release #1.0, Version #1.25
22 : CURRENT APPLICATION DATA:
23 : APPLICATION NUMBER: US/08/201.118
24 : FILING DATE: 22-FEB-1994
25 : CLASSIFICATION: 435
26 : PRIOR APPLICATION DATA:
27 : APPLICATION NUMBER: US 07/864,962
28 : FILING DATE: 09-APR-1992
29 : ATTORNEY/AGENT INFORMATION:
30 : NAME: ALBERSGREN, JON
31 : REGISTRATION NUMBER: 505
32 : REFERENCE/DOCKET NUMBER: 15280-192-1
33 : TELECOMMUNICATION INFORMATION:
34 : TELEPHONE: (415) 326-2400
35 : TELEFAX: (415) 326-2422
36 : INFORMATION FOR SEQ ID NO: 6:
37 : SEQUENCE CHARACTERISTICS:
38 : LENGTH: 2009 base pairs
39 : TYPE: nucleic acid
40 : STRANDEDNESS: Single
41 : TOPOLOGY: linear
42 : MOLECULE TYPE: CDNA
43 : US-08-201-118-6
44
45 Query Match 52.01: Score 168: DB 1: Length 2009:
46 Best Local Similarity 92.21:
47 Matches 177: Conservative 0: Mismatches 15: Indels 0: Gaps 0:

```

Search completed: April 19, 2002, 08:10:43
Job time: 2943 sec

XX Polynucleotides comprising sequences from malate decarboxylase
 PF enzyme-related biallelic markers used for genotyping
 PS
 XX Claim 13; Page 281: 67pp; English.
 XX Sequences AAH51110-AAH51593 represent human DNA fragments which contain
 CC biallelic markers. The sequences are related to various human genes
 CC including microsomal glutathione S-transferase II (MGSTII), malate
 CC decarboxylase enzyme (MDL1/MDL2), cytochrome P450, glutathione
 CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),
 CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate
 CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
 CC diphosphate glucosyl transferase (UGT2). Each of these sequences
 CC contains a biallelic marker/polymorphism, which is represented in the
 CC sequence by a "tag" or "tag-a" or "tag-b" or "tag-c". The tag is a
 CC marker containing sequences are related are involved in drug metabolism.
 CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
 CC MGSTII gene and four alternative MGSTII cDNA sequences. AAH62905-AAH62906
 CC are MGSTII gene products. PCR primers AAH51599 and AAH51600 are used in
 CC an example for the amplification of human genomic DNA fragments. The
 CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
 CC a biological sample. The method is used to determine the frequency in
 CC population of an allele of a DME- or MGSTII-related biallelic marker and
 CC to determine the frequency in a population of a DME- or MGSTII-related
 CC treatment. The method is also used to detect a relation between allele
 CC and phenotype, and to detect association between haplotype and phenotype.
 CC The polynucleotides are used, in hybridization assays, sequencing assays
 CC or allele specific amplification assays. The method can be used to
 CC determine whether an individual suffers or is at risk of developing
 CC asthma or is at risk of developing hepatotoxicity on treatment with
 CC ziluton.
 XX
 XX Sequences 1001 BP: 270 A; 166 C; 178 G; 384 T; 3 other;

Query Match 69.0%; Score 223; DB 21; Length 1001;
 Best Local Similarity 83.0%; Pred. No. 5.3e-62;
 Matches 253; Conservative 1; Mismatches 51; Indels 0; Gaps 0;
 Oy 2 cctgaattgtacacaaatgtgccattttttctctttccatcagttttacttgt 61
 Db 197 ccatgaattgtacacaaatgtccatatactctgtttccatcagttttacttgt 256
 Oy 62 ctatcagcttaagtcgaagagagattgacgtgtgtgtgtgtgtgtgtgtgtgt 121
 Db 257 ctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 316
 Oy 122 catgcagagcagagcagcagcagcagcagcagcagcagcagcagcagcagcag 181
 Db 317 catgcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 376
 Oy 182 ccttgaccttctccacacagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 241
 Db 377 cagtgagcgttgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 436
 Oy 242 ctatctctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagc 301
 Db 437 ctactctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 496
 Oy 302 ccaaa 306
 Db 497 tccaw 501

RESULTS 3
 AAH11378 standard; cDNA; 1854 BP.
 XX AAH11378;
 XX 09-SEP-1996 (first entry)
 DT

XX Cytochrome P450 2C9 clone 25 coding sequence.
 DE
 XX Cytochrome P450 2C19; human; liver; PCR; primer; detection: CYP2C19;
 XX stereospecific S-mephenytoin 4'-hydroxylase activity; polymorphism; ss.
 XX Homo sapiens.
 OS
 XX Key Location/Qualifiers
 XX CDS 13..1485
 XX /tag- a
 XX /product- Cytochrome P450 C9 clone 25
 XX misc_difference 1087
 XX /tag- b
 XX /note- "Variable position"
 XX misc_difference 1087
 XX /tag- c
 XX /note- "Variable position"
 XX W09530766-A1.
 XX 16-NOV-1995. 95MO-US05744.
 XX 08-MAY-1995; 94US-0238821.
 XX 06-MAY-1994; 94US-0238821.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX De Morals SMF, Goldstein JA, Romkes-Sparks M;
 XX WPI: 1996-077257/08.
 XX P-PSDB: AAR89862.
 XX New isolated cytochrome P450 2C subfamily member - used for
 XX identifying drugs metabolized by S-mephenytoin 4'-hydroxylase
 XX activity and to develop other screening assays
 XX Example 2; Page 95-96; 169pp; English.
 XX The sequences given in AAT11378-81 encode allelic variants of cytochrome
 CC P450 2C9. The majority of clones isolated from liver S33 coded for 2C9.
 CC Of the 50 clones encoding 2C9, only two allelic variants were found. 39
 CC of the 2C9 clones were identical with clone 65, and 11 were identical
 CC with clone 25. Clones 25 and 65 are identical in the 5' and
 CC 3'-noncoding regions but contained 2 single base changes at positions
 CC 1075 and 1425. One of these base changes was conservative but the other
 CC resulted in one amino acid difference at position 359 (Ile to Leu). The
 CC position 1154, which results in a single amino acid change, Thr to Met
 CC at position 385. Clone 29c has a very long, 198 bp, 5'-noncoding region
 CC and a polyadenylation signal 21 bases from the poly-A tail. Clone 6b
 CC has an unusually long 3'-noncoding region containing three possible
 CC polyadenylation signals with no poly-A tail. The differences in the 3'
 CC non-coding regions could represent alternate splicing, allelic
 CC variants, or possibly separate genes. These clones are designated as
 CC allelic variants of 2C18 because they differ by only one base in the
 CC coding region. They are most similar to 2C9 (82% amino acid homology)
 CC and 2C19 (81% amino acid homology).
 XX Sequence 1854 BP: 513 A; 424 C; 381 G; 536 T; 0 other;

Query Match 59.4%; Score 192; DB 17; Length 1854;
 Best Local Similarity 100.0%; Pred. No. 7.3e-52;
 Matches 192; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 67 cagctaaagtccaggaagattgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 126
 Db 971 cagctaaagtccaggaagattgacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1030
 Oy 127 aagacagagaccatgcctacacatgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 186

DB	1031	aaagacaggagccatcgcctcctacacagatgctgctggtcacagagtcagagataccttg	1090
OT	187	acctctctcccacagcgtgcgccatgcagtcagctatgacattcaattcagaacatc	246
DB	1091	acctctctcccacagcgtgcgccatgcagtcagctgacattcaattcagaacatc	1150
OT	247	tcattctcccaag	258
DB	1151	tcattctcccaag	1162
RESULT 4			
AAV44154			
ID	AAV44154	standard; cDNA; 1854 BP.	
AC	AAV44154;		
OT	06-OCT-1998	(first entry)	
DE	Human cytochrome P450 2C9 clone 25 cDNA.		
KW	Cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;		
KW	identification; autagenic; carcinogenic; cytotoxic; haemoprotein;		
KW	zenobiotic; environmental pollutant; ss.		
XX	Howe sapiens.		
XX	US5786191-A.		
XX	PR		
XX	28-JUL-1998.		
PD			
XX	22-FEB-1994; 94US-0201118.		
PF			
XX	22-FEB-1994; 94US-0201118.		
PR			
XX	09-APR-1992; 92US-0864962.		
PR			
XX	(GOLD/) GOLDSTEIN J A.		
PA	(HOWE) HOWES-SPARKS M.		
PA			
PI	Goldstein JA, Howes-sparks M;		
PI	WPI: 1998-436528/37.		
DR			
XX	Screening for drugs metabolised by cytochrome P450 - for identifying		
PT	autagenic, carcinogenic, or cytotoxic compounds		
PT			
PS	Example 2; Column 39-42; 63pp; English.		
PS			
CC	This sequence encodes a cytochrome P450 2C9 polypeptide isolated from		
CC	human clone 25. This polypeptide is used in a method to screen for a drug		
CC	that is metabolised by a cytochrome P450 having S-mephenytoin		
CC	4'-hydroxylase activity. The protein can also be used to identify a		
CC	autagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a		
CC	subclass of haemoprotein enzymes capable of metabolising xenobiotics		
CC	such as drugs, environmental pollutants as well as		
CC	endobiotics such as steroids, fatty acids and prostaglandins.		
XX	Sequence 1854 BP; 513 A; 424 C; 381 G; 536 T; 0 other.		

	Query Match	59.4%	Score 192	DB 19	Length 1854
	Best Local Similarity	100.0%	Pos No 73-72		
	Matches 192	Conservative 0	Mismatches 0	Indels 0	Gaps 0
07	67	cagctcaagtccagaagaagatgaacgtgtattgacgaacaggagccctgcgacgc	126		
0b	971	cagctcaagtccagaagaagatgaacgtgtattgacgaacaggagccctgcgacgc	1030		
07	127	aagcagagggccacagtcctcacaagatctctgtgtgacgaggtccagagatccttg	186		
0b	1031	aagcagagggccacagtcctcacaagatctctgtgtgacgaggtccagagatccttg	1090		
07	187	acctctctccacacagctgcgccacagcagtcgactctgtgacatctaaatcagaaacatc	246		

```

Db      1091 acctctccaccagcagctcccacgagtgaccctgacattaaatcagaactatc 1150
Oy      247  taattccaagg 258
Db      1151 taattccaagg 1162

RESULT 5
AXI19916
ID      AAXI19916 standard; DNA; 1356 BP.
XX      AAXI19916;
XX      DT
XX      11-JUN-1999 (first entry)
DE      Bacterial and mammalian chimeric cytochrome P450 protein encoding DNA.
KW      Bacterial; mammalian; cytochrome P450; chimeric; fusion protein;
KW      oxidise; hydrocarbon; carbon-hydrogen bond; hydroxylating;
KW      bioremediation; environmental pollutant; ss.
XX      Synthetic.
XX      XS
XX      XS
PN      HQ9508812-A1.
PD      25-FEB-1999.
PF      17-AUG-1998: 98wo-US16979.
PR      20-AUG-1997: 9705-0056754.
XX      (UTRP ) UNIV ROCHESTER.
PI      Jones JP, Shimoji M;
XX      WPI: 1999-190131/16.
XX      P-PSDB: AAY04126.
XX      New P450 fusion proteins - comprising a portion of a bacterial
XX      cytochrome P450 protein and a portion of a mammalian cytochrome P450
XX      protein
XX      Claim 10; Page 5-6; 51pp; English.
XX      The present sequence encodes a fusion proteins comprising a portion of a
XX      bacterial cytochrome P450 protein and also a portion of a mammalian
XX      cytochrome P450 protein. The fusion protein can oxidise hydrocarbons or
XX      any compound having a carbon-hydrogen bond. The fusion protein can be
XX      used for hydroxylating a compound to be oxidised. It can also be used in
XX      the remediation of an environmental pollutant. Since the fusion
XX      protein is soluble in water, it can be used for bioremediation by x-ray
XX      crystallography for designing functional proteins that can be readily
XX      expressed in soil bacteria to facilitate bioremediation.
XX      Sequence 1356 BP; 349 A; 386 C; 323 G; 298 T; 0 other;
XX      SO

Query Match 58.94; Score 190.4; DB 20; Length 1356;
XX      Accession: U00534;
XX      Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy      67  cagctaagaatccaggagaagattgaacgtgtgattgacagaacacggccctgcatgc 126
Db      842  cagctaagaatccaggagaagattgaacgtgtgattgacagaacacggccctgcatgc 901
Oy      127  aagacagagagccatgcctctacacagatgctgtgtgcacgagctccagagataccttg 186
Db      192  aagacagagagccatgcctctacacagatgctgtgtgcacgagctccagagataccttg 961
Oy      187  acctctcccaaccagcctgcccatgcagctgacccgtgacatttaattcagaactatc 246
Db      962  acctctcccaaccagcctgcccatgcagctgacccgtgacatttaattcagaactatc 1021

```

Query Match	58.9%	Score 190.4	DB 20	Length 1356
Best Local Similarity	95.3%	Pred. No. 2.1e-31		
Matches 191	Conservative	0 Mismatches	1 Indels	0 Gaps
67	cagctcaaaattccaggagaattgagcgtgattgtgcagaacacgcgacctgcgacg	126		
842	cagctcaaaattccaggagaattgagcgtgattgtgcagaacacgcgacctgcgacg	901		
127	agacgaggagccacgctctacacagatgctgtgtgcacagctcccgagataccctg	186		
902	agacgaggagccacgctctacacagatgctgtgtgcacagctcccgagataccctg	961		
187	actctctcccaacacgctgcccacatgcgctgcgacatgaattcagaacatc	246		
962	actctctcccaacacgctgcccacatgcgctgcgacatgaattcagaacatc	1021		

QY 247 tcaattccaagg 258
 DB 1022 tttttttttttt 1033
 RESULT 6.
 ID AQ087715 standard: cDNA; 1473 BP.
 AC AQ087715;
 XX 10-MAY-1995 (first entry)
 DE Human cytochrome P450 molecular species 2C9 cDNA.
 KW Human cytochrome P450; amplification; PCR; primer: expression vector;
 KW yeast NADPH-P450 reductase; safety: fusion protein; metabolite:
 KW carcinogen; mutagen; liver metabolism; ds.
 XX Homo sapiens.
 OS EF644267-A.
 PN 22-MAR-1995.
 XX 20-JUL-1994; 94EP-0111298.
 XX 21-JUL-1993; 93JP-0180246.
 XX 20-JUL-1993; 93JP-0201120.
 XX 30-JUL-1993; 93JP-0208279.
 XX (BATA) HAYASHI K.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 PI Bayashi K, Kaneko H, Konal K, Nakatsuka I, Sakaki T;
 PI Tabuchi T;
 WPI: 1995-116991/16.
 DR P-PSDB; AAR72361.
 XX Evaluation of safety of a chemical cpd. - using recombinant yeast
 XX expressing human cytochrome P450 and a yeast NADPH-P450 reductase
 XX Examples; Page 23-25; 124pp: English.
 XX The nucleotide sequence of the cDNA encoding the human cytochrome P450
 XX species 2C9. The gene encodes a protein of 490 amino acids. The gene
 XX was amplified by PCR using the primers AQ087735-8. The product was
 XX cloned into the yeast expression vectors pAH5N or pARRR to produce the
 XX ve-rms p2C9 for the expression of the cytochrome P450 alone or p2C9R
 XX for co-expression with the yeast NADPH-P450 reductase.
 XX The vectors were used in a method for evaluating the safety of a chemical
 XX compound by reacting the chemical compound with recombinantly produced
 XX human cytochrome P450 molecular species 1A2 (AA087714), 2C9, 2E1
 XX (AQ087718-32) and yeast NADPH-P450 reductase. The resulting protein
 XX or in cell extracts and analysing the resulting metabolite to assess the
 XX safety of the chemical compound. The method is useful for determining
 XX whether the chemical compound, or its metabolite, will be converted into
 XX a carcinogenic or mutagenic form through metabolism in the liver.
 XX Sequence 1473 BP: 413 A: 345 C: 319 G: 396 T: 0 other;
 XX Query Match 58.9%; Score 190.4; DB 16: Length 1473;
 XX Best Local Similarity 99.5%; Pred. No. 2,2e-51;
 XX Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 67 cagctaaagtcaggagagattgaacgctgattggcagacacccaggccctctcatc 126
 DB 959 cagctaaagtcaggagagattgaacgctgattggcagacacccaggccctctcatc 1018

OY 67 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 959 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 1018
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagccacagataacttcg 186
 DB 1019 aagacagagccacatgcctccacacagatgctgtgattgacagagccacagataacttcg 1078
 OY 187 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 246
 DB 1079 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 1138
 OY 247 tcaattcccaagg 258
 DB 1139 tcaattcccaagg 1150

RESULT 8

AAT17404
 ID AAT17404 standard; cDNA; 1473 BP.
 AC AAT17404;
 DT 01-AUG-1996 (first entry)
 DE Human derived cytochrome P4502C9 cDNA.
 EN Human derived cytochrome: P4502C9; commercial cDNA library; yeast;
 KW transfection; recombinant production; expression vector; mammal;
 KW immunisation; sensitisation; antibody; determination; detection;
 KW non-cross reactive; ds.
 OS Homo sapiens.
 FX Key Location/Qualifiers
 FT CDS 1..1473
 FT /tag= a
 XX JP08027196-A.
 XX 30-JAN-1996.
 XX 13-JUL-1994; 94JP-0161551.
 XX 13-JUL-1994; 94JP-0161551.
 XX (SDMO) SUMITOMO CHEM CO LTD.
 XX WPI; 1996-136337/14.
 XX P-PSDB; AAR81465.
 XX Antibody recognising human derived cytochrome P4502C9 - allows
 XX specific detection of cytochrome P450 species in humans
 XX Example 1: Pages 11-13; 13pp; Japanese.
 XX The present sequence encodes the human derived cytochrome (HDC)
 XX P4502C9 which was subcloned into a pUC19 vector. The recombinant
 XX were transfected with an expression vector into a yeast host,
 XX cultured and then disrupted to give a microsomal fraction. The
 XX HDC was purified from the fraction, and used to immunise and
 XX sensitise a mammal. Blood was drawn from the mammal, and an
 XX anti-HDC antibody isolated. The antibody obtd. recognises HDC
 XX P4502C9, partic. at a serum dilution rate of 1:10000, and is
 XX substantially without cross reaction to other HDC P450 spp..

Query Match 58.9%; Score 190.4; DB 17; Length 1473;
 Best Local Similarity 99.5%; Pred. No. 2.2e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 67 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 959 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 1018
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagccacagataacttcg 186
 DB 1019 aagacagagccacatgcctccacacagatgctgtgattgacagagccacagataacttcg 1078
 OY 187 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 246
 DB 1079 acctctcccccacagcctgcccacatgcagctgacattgaattcagaactatc 1138
 OY 247 tcaattcccaagg 258
 DB 1139 tcaattcccaagg 1150

RESULT 9

AA003599
 ID AA003599 standard; DNA; 1818 BP.
 AC AA003599;
 DT 03-SEP-1990 (first entry)
 DE Human liver cytochrome P-450 encoding gene.
 EN P-450; cytochrome; Saccharomyces cerevisiae;
 KW liver disorders; ds.
 OS Homo sapiens.
 FX Key Location/Qualifiers
 FT CDS 1..1443
 FT /tag= a
 XX JP02072879-A.
 XX 13-MAR-1990.
 XX 09-SEP-1988; 88JP-0225955.
 XX 09-SEP-1988; 88JP-0225955.
 XX (AGEN) AGENCY OF IND SCI TECH.
 XX WPI; 1990-121045/16.
 XX P-PSDB; AAR04043.
 XX Recombinant plasmid, for yeast for liver treatment -
 XX comprises human liver cytochrome P-450HP gene obtd. from yeast.
 XX etc.
 XX Disclosure: Fig 4; 9pp; Japanese.
 XX The P-450 produced from plasmid pHPA6 in a yeast expression system
 XX preferably Saccharomyces AD22, can be used to treat human liver
 XX disorders, oxidising various chemical substances.
 XX Sequence 1818 BP; 516 A; 412 C; 376 G; 514 T; 0 other;

Query Match 58.9%; Score 190.4; DB 11; Length 1818;
 Best Local Similarity 99.5%; Pred. No. 2.4e-51;
 Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 67 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 126
 DB 932 cagctaaagtcaggagagagattgaacgtgtgattggcagaacccgagccctgcatgc 991
 OY 127 aagacagagccacatgcctccacacagatgctgtgattgacagagccacagataacttcg 186
 DB 1139 tcaattcccaagg 1150

CC variants, or possibly separate genes. These clones are designated as
CC all as cytochrome P450 2C9 because they differ by only one base in the
CC coding region. They are most similar to 2C9 (82% amino acid homology)
CC and 2C19 (81% amino acid homology).

XX Sequence 1852 BP: 514 A: 424 C: 380 G: 534 T: 0 other:

Query Match 58.9% Score 190.4; DB 17: Length 1852;
Best Local Similarity 99.5%; Pred. No. 2.4e-51;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 67 cagctaaagtcacagagagattgaacgtgtgattgacagaaacccgagccctgcatgc 126
Db 969 cagctaaagtcacagagagattgaacgtgtgattgacagaaacccgagccctgcatgc 1028
Oy 127 agacagagagccatgcctcacacagatgctgtgtgacagaggtccagataacttg 186
Db 1029 agacagagagccatgcctcacacagatgctgtgtgacagaggtccagataacttg 1088
Oy 187 accttctccacacagctccctccatgcagtgacgtgacgttaattcagaactatc 246
Db 1089 accttctccacacagctccctccatgcagtgacgtgacgttaattcagaactatc 1148
Oy 247 tcatttcccaagg 258
Db 1149 tcatttcccaagg 1160

RESULT 12

AAH51162 standard; cDNA; 1852 BP.

AC AAH51162:

XX

XX 29-AUG-2001 (first entry)

XX

XX Human CYP2C9 related DNA containing a biallelic polymorphism SEQ ID 53.

XX

XX Human: biallelic marker: single nucleotide polymorphism: SNP: MGST11;

XX cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;

XX identification; mutagenic; carcinogenic; cytotoxic; haemoprotein;

XX xenobiotic; environmental pollutant; ss.

XX

XX Homo sapiens.

XX

XX US5786191-A.

XX

XX 28-JUL-1998.

XX

XX 22-FEB-1994; 94US-0201118.

XX

XX 22-FEB-1994; 94US-0201118.

XX

XX 09-APR-1992; 92US-0864962.

XX

XX (GOLD)/ GOLDSTEIN J A.

XX

XX (R&M)/ ROKES-SPARKS M.

XX

XX Goldstein JA, Rokes-sparks M;

XX

XX WPI: 1998-436528/37.

XX

XX Screening for drugs metabolised by cytochrome P450 - for identifying

XX mutagenic, carcinogenic, or cytotoxic compounds

XX

XX Example 2; Column 53-56; 63pp: English.

XX

XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated

XX from clone 85. This polypeptide is a member of the cytochrome P450

XX family, which is a large family of haemoproteins. The protein can also be used to identify a

XX 4'-hydroxylase activity. The protein can also be used to identify a

XX autagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a

XX large family of haemoprotein enzymes capable of metabolising xenobiotics

XX such as drugs carcinogens and environmental pollutants as well as

CC endobiotics such as steroids, fatty acids and prostaglandins.

XX Sequence 1852 BP: 514 A: 424 C: 380 G: 534 T: 0 other:

Query Match 58.9% Score 190.4; DB 19: Length 1852;

Best Local Similarity 99.5%; Pred. No. 2.4e-51;

Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaagtcacagagagattgaacgtgtgattgacagaaacccgagccctgcatgc 126

Db 969 cagctaaagtcacagagagattgaacgtgtgattgacagaaacccgagccctgcatgc 1028

Oy 127 agacagagagccatgcctcacacagatgctgtgtgacagaggtccagataacttg 186

Db 1029 agacagagagccatgcctcacacagatgctgtgtgacagaggtccagataacttg 1088

Oy 187 accttctccacacagctccctccatgcagtgacgtgacgttaattcagaactatc 246

Db 1089 accttctccacacagctccctccatgcagtgacgtgacgttaattcagaactatc 1148

Oy 247 tcatttcccaagg 258

Db 1149 tcatttcccaagg 1160

RESULT 13

AAH51162 standard; DNA; 1001 BP.

AC AAH51162:

XX

XX 29-AUG-2001 (first entry)

XX

XX Human CYP2C8 related DNA containing a biallelic polymorphism SEQ ID 53.

XX

XX Human: biallelic marker: single nucleotide polymorphism: SNP: MGST11;

XX cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;

XX identification; mutagenic; carcinogenic; cytotoxic; haemoprotein;

XX xenobiotic; environmental pollutant; ss.

XX

XX Homo sapiens.

XX

XX WO200058508-A2.

XX

XX 05-OCT-2000.

XX

XX 24-MAR-2000; 2000WO-IB00403.

XX

XX 25-MAR-1999; 99US-0126269.

XX

XX 30-APR-1999; 99US-0131961.

XX

XX (GSE) GENSET.

XX

XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;

XX

XX WPI: 2000-638353/61.

XX

XX Polynucleotides comprising sequences from malate decarboxylase

XX enzyme-related biallelic markers used for genotyping -

XX

XX Claim 13; Page 283; 673pp: English.

XX

XX Sequences AAH51110-AAH51593 represent human DNA fragments which contain

XX biallelic markers. The sequences are related to various human genes

XX including microsomal glutathione S-transferase II (MGSTII), malate

XX decarboxylase enzyme (DMEI/MEI), cytochrome P450, glutathione

XX reductase/synthase (GSHR/GSHS), flavin-containing monooxygenases (FMO),

XX gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 6-phosphate

67 cagctaaagtcacggaagagattgaacgtgtgattggcagaaacaggagccccctgcattgc 176

TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Activation of Gene Expression
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave. Cleveland, OH 44115, USA
Phone: 216 361 9000
Fax: 216 361 9596
Email: scain@atersys.com
FEATURES High quality sequence stop: 433.
Location/Qualifiers
1. .901
/organism="Homo sapiens"

/cell_line="H1080"
/note="See Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression" Nature Biotechnology, in press. Note that even though the cell type indicated is H1080, since a random activation

```

BASE COUNT      230 a      241 c      199 g      231 t
ORIGIN
Query Match      58.94;      Score 190.4;      DB 11;      Length 901;
Best Local Similarity 99.54;      Pred. No. 5,2e+41;
Matches 191;      Conservative 0;      Mismatches 1;      Indels 0;      Gaps 0;

67  cagctaaagtcgaagagagattgaactgtattgtcagaacacgcgcctcgatgc 126
      |||||
148  CAGCTAAAGTCCGAAGAGAGATTGAACGTGTATTGGCAGAAACCGGCGCCCTCTATGC 207
      |||||

```

	208	187	247	328	RESULT	8	EST	21-APR-2001	
Db	AAGACAGAGCCACATGCCCTACACAGATCCTGTGTGTCAGAGGTCGACAGATACATGT	acctctctcccccagcctgcgccatgcagtcagctctgacatttaattcagaanaactatc	tcattctcccaag	TCATTCCCAGG	LOCUS	BG205076	907 bp	mRNA	
Db					DEFINITION	R574495	Athersys RAGE Library	EST	
Db					ACCESSION	BG205076			
Oy					VERSION	BG205076.1	G1:13726763		
Db					KEYWORDS	EST.			

REFERENCE
Mammalia: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
1 Mammalia: Eutheria: Primates: Catarrhini: Homidae: Homo.
1 (bases 1 to 907)
Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Dahl, T., Thorncorn, M., Ramachandran, R., Whittington,
Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Reys, R., Smith, E.,
Veloso, N., Hess, J., Cottrill, K., J. Lo, K., Offenbacher, J., Danzig, J.,
and Ducar, M. Genome-wide Protein Expression Libraries using Random
Mutations of Gene Expression
Nat Biotechnol. 19 (5): 440 (2001) In press
Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave. Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com

High quality sequence stop: 354.
Location/Qualifiers
1. 907
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 226 a 234 c 193 g 254 t
ORIGIN

Query Match 58.9% Score 190.4; DB 11; Length 907;
Best Local Similarity 99.5%; Pred. No. 5,2e-41;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaagtcaggagagattgacgctggtggtgacgaacccgagccctgcacgc 126
|||||
Db 145 CAGCTAAAGTCAGGAGAGATTGACGCTGTTGTCGACGAGTCCAGAGATACATG 204
|||||

Oy 127 aagcagagccacatgccctacacagatgctggtgacgaggtccagagataccttg 186
|||||
Db 205 AAGCAGAGCCACATGCCCTACACAGATGCTGTTGTCGACGAGTCCAGAGATACATG 264
|||||

Oy 187 acccttcctccacagcctccctccctgacgctgacattaaattcagaactatc 246
|||||
Db 265 AACTTCCTCCACAGCCTCCCTCCCTGACCTGTGACATTAAATTCAGAACTATC 324
|||||

Oy 247 tcattcccaagg 258
|||||
Db 325 TCATTCCCAAGG 336
|||||

RESULT 9
LOCUS BGI98889 909 bp mRNA EST 21-APR-2001
DEFINITION RST18160 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BGI98889
VERSION BGI98889.1 GI:13720576
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 909)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E., Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 532.
Location/Qualifiers
1. 909
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 113 a 104 c 90 g 81 t
ORIGIN

Query Match 58.5% Score 188.8; DB 11; Length 388;

BASE COUNT 237 a 240 c 181 g 250 t
ORIGIN

Query Match 58.9% Score 190.4; DB 11; Length 909;
Best Local Similarity 99.5%; Pred. No. 5,2e-41;
Matches 191; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 67 cagctaaagtcaggagagattgacgctggtggtgacgaacccgagccctgcacgc 126
|||||
Db 147 CAGCTAAAGTCAGGAGAGATTGACGCTGTTGTCGACGAGTCCAGAGATACATG 206
|||||

Oy 127 aagcagagccacatgccctacacagatgctggtgacgaggtccagagataccttg 186
|||||
Db 207 AAGCAGAGCCACATGCCCTACACAGATGCTGTTGTCGACGAGTCCAGAGATACATG 266
|||||

Oy 187 acccttcctccacagcctccctccctgacgctgacattaaattcagaactatc 246
|||||
Db 267 AACTTCCTCCACAGCCTCCCTCCCTGACCTGTGACATTAAATTCAGAACTATC 326
|||||

Oy 247 tcattcccaagg 258
|||||
Db 327 TCATTCCCAAGG 338
|||||

RESULT 10
LOCUS BGI84407 388 bp mRNA EST 21-APR-2001
DEFINITION RST3332 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BGI84407
VERSION BGI84407.1 GI:13706094
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 388)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J., Lerner, L., Krashinsky, D., McElligott, K., Clark, S., Mays, R., Smith, E., Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J., and Ducar, M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 338.
Location/Qualifiers
1. 388
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 113 a 104 c 90 g 81 t
ORIGIN

Query Match 58.5% Score 188.8; DB 11; Length 388;

```

Best Local Similarity 99.0%; Pred. No. 1.3e-40; Mismatches 0; Indels 0; Gaps 0;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaatccagagagattgacgtgattgagcagaacacgagccctgcatgc 126
Db 162 CAGCTAAAGTCCAGAGAGATTGACGTGTGATTGGCAGAAACCGAGCCCTGCATGC 241

Oy 127 aagcagagagacatgctccacacagatgctggtgacgaggtccgagataccttg 186
Db 242 AAGAGAGAGAGACATGCTCCACACAGATGCTGTGTGGCAGAGGTCCAGAGATACATTC 301

Oy 187 acctctccacacagctccacacagatgctggtgacgaggtccgagataccttg 246
Db 302 ACCTCTCCACACAGCTCCACACAGATGCTGTGTGGCAGAGGTCCAGAGATACATTC 361

Oy 247 tcattcccaagg 258
Db 362 TCATTCCCAAGG 373

RESULT 11
LOCUS AV651261 620 bp mRNA EST 07-SEP-2000
DEFINITION AV651261 GLC Homo sapiens cDNA clone GLC126 3', mRNA sequence.
ACCESSION AV651261
VERSION AV651261.1 GI:9872275
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 620)
AUTHORS Qian,B., Wu,T., Huang,O., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
, Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Ou,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens cDNA clone
Unpublished (2000)
Contact: Zengqiang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shouling Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801913(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..620
/organism="Homo sapiens"
/clone="GLC126"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"
BASE COUNT 205 a 141 c 112 g 162 t
ORIGIN
Query Match 58.5%; Score 188.8; DB 10; Length 620;
Best Local Similarity 99.0%; Pred. No. 1.3e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaatccagagagattgacgtgattgagcagaacacgagccctgcatgc 126
Db 379 CAGCTAAAGTCCAGAGAGATTGACGTGTGATTGGCAGAAACCGAGCCCTGCATGC 438

Oy 127 aagcagagagacatgctccacacagatgctggtgacgaggtccgagataccttg 186
Db 439 AAGAGAGAGAGACATGCTCCACACAGATGCTGTGTGGCAGAGGTCCAGAGATACATTC 498

Best Local Similarity 99.0%; Score 188.8; DB 10; Length 661;
Query Match 58.5%; Score 188.8; DB 10; Length 661;
Best Local Similarity 99.0%; Pred. No. 1.4e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 67 cagctaaatccagagagattgacgtgattgagcagaacacgagccctgcatgc 126
Db 52 CAGCTAAAGTCCAGAGAGATTGACGTGTGATTGGCAGAAACCGAGCCCTGCATGC 111

Oy 127 aagcagagagacatgctccacacagatgctggtgacgaggtccgagataccttg 186
Db 112 AAGAGAGAGACATGCTCCACACAGATGCTGTGTGGCAGAGGTCCAGAGATACATTC 171

Oy 187 acctctccacacagctccacacagatgctggtgacgaggtccgagataccttg 246
Db 172 ACCTCTCCACACAGCTCCACACAGATGCTGTGTGGCAGAGGTCCAGAGATACATTC 231

Oy 247 tcattcccaagg 258
Db 232 TCATTCCCAAGG 243

RESULT 13
BG217233

```

LOCUS BG1217233 802 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG1217233
 VERSION BG1217233.1 GI:13743254
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 802)
 AUTHORS Barrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9500
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 473.
 Location/Qualifiers
 1..802
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression', the
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 200 a 217 c 171 g 214 t
 ORIGIN
 Query Match 58.5% Score 188.8; DB 11; Length 802;
 Best Local Similarity 99.0%; Pred. No. 1.4e-40;
 Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
 Oy 67 cagctaagtcacgaagagattgaacgtgtgattggcgaacacggagccctgcacg 126
 Db 145 CAGCTAAAGTCCGAGGAGATTGAACGTGTGATTGGCAGAACCGGAGCCCTGCATGC 204
 Oy 127 aagcagagacacacgtgccctcacacagatgctgtgacgaggtccagagataccttg 186
 Db 205 AAGCAGAGGAGCCACATGCCCTCACACAGATGCTGTGTGGCGGAGTCCGAGATACATTC 264
 Oy 187 accttctcccccacagctgccctcacacagatgctgtgacgaggtccagagataccttg 186
 Db 265 ACCTTCTCCCCACAGCTGCCCTCACACAGATGCTGTGTGGCGGAGTCCGAGATACATTC 324
 Oy 247 taattcccaagg 258
 Db 325 TCATTCCTCAAGG 336
 RESULT 14
 LOCUS BG196822 876 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG196822
 VERSION BG196822.1 GI:13718509
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 876)
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
 COMMENT Contact: Scott J. Cain
 Athersys, Inc.
 3201 Carnegie Ave, Cleveland, OH 44115, USA
 Tel: 216 431 9500
 Fax: 216 361 9596
 Email: scain@atersys.com
 High quality sequence stop: 456.
 Location/Qualifiers
 1..876
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Athersys RAGE Library"
 /cell_line="HT1080"
 /note="See 'Creation of Genome-wide Protein Expression
 Libraries using Random Activation of Gene Expression', the
 Nature Biotechnology, in press. Note that even though the
 cell type indicated is HT1080, since a random activation
 method was used, these sequence tags are not necessarily
 expressed in HT1080 under normal circumstances."
 BASE COUNT 222 a 222 c 194 g 237 t 1 others
 ORIGIN
 Query Match 58.5% Score 188.8; DB 11; Length 876;
 Best Local Similarity 99.0%; Pred. No. 1.4e-40;
 Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
 Oy 67 cagctaagtcacgaagagattgaacgtgtgattggcgaacacggagccctgcacg 126
 Db 144 CAGCTAAAGTCCGAGGAGATTGAACGTGTGATTGGCAGAACCGGAGCCCTGCATGC 203
 Oy 127 aagcagagacacacgtgccctcacacagatgctgtgacgaggtccagagataccttg 186
 Db 204 AAGCAGAGGAGCCACATGCCCTCACACAGATGCTGTGTGGCGGAGTCCGAGATACATTC 263
 Oy 187 accttctcccccacagctgccctcacacagatgctgtgacgaggtccagagataccttg 246
 Db 264 ACCTTCTCCCCACAGCTGCCCTCACACAGATGCTGTGTGGCGGAGTCCGAGATACATTC 323
 Oy 247 taattcccaagg 258
 Db 324 TCATTCCTCAAGG 335
 RESULT 15
 LOCUS BG197837 896 bp mRNA EST 21-APR-2001
 DEFINITION Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
 ACCESSION BG197837
 VERSION BG197837.1 GI:13719524
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 896)
 AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
 Cain, S., Dahl, T., Thornton, M., Ramachandran, R., Whittington, J.,
 Lerner, L., Krashoc, D., McElligott, K., Clark, S., Mays, R., Smith, E.,
 Veloso, N., Hess, J., Cothren, K., Lo, K., Offenbacher, J., Danzig, J.,
 and Ducar, M.
 TITLE Creation of Genome-wide Protein Expression Libraries using Random
 Activation of Gene Expression
 JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press

COMMENT

Contact: Scott J. Cain
Nucleus, Inc.
3200 West 10th Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@nucleusys.com
High quality sequence stop: 321.
Location/Qualifiers

FEATURES

source
1. .896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HT1080"
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT

221 a 251 c 181 g 242 t 1 others

ORIGIN

Query Match 58.5%; Score 188.8; DB 11; Length 896;
Best Local Similarity 99.0%; Pred. No. 1.4e-40;
Matches 190; Conservative 0; Mismatches 2; Indels 0; Gaps 0:
QY 67 cagctaaagtcacgaagagattgaacgtgtgattggcagaacccgagccctcgtatgc 126
DB 114 CAGCTAAGTCCGAGAGAGATGAGCGTGTGATGGCAGAACCCGAGCCCTCGCATGC 173
QY 127 aacacagagacacatgcctacacagatgctgtgtatgcacagatccagatacccttg 186
DB 174 AAGACAGACCCACATGCCCTACACAGATGCTGTGTGTGTCACGAGGTCCAGAGATACATIG 233
QY 187 accttctcccccacagcctgcgccatgcagtgacacctgtgacattaaattccagaactatc 246
DB 234 AACTTCTCCCCACAGCCTGCCCATGCTGACCTGTGACATTAAATTCAGAAACTATC 293
QY 247 coattccaaag 258
DB 294 TATTCCTCAAG 305

Search completed: April 19, 2002, 08:07:52
Job time: 2772 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK nucleic - nucleic search, using sv model

Run on: April 19, 2002, 11:06:57 : Search time 232.77 Seconds
(without alignments)
163.811 Million cell updates/sec

Title: US-09-763-292-3
Perfect score: 33
Sequence: 1 cccggaattactacaa.....aaatcataatcatatttt 323
Scoring table: IDENTITY.MUC
Gapop 10.0, Gapext 1.0
Searched: 78474 seqs, 5898373 residues

Total number of hits satisfying chosen parameters: 156948

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents_NA_New.*

- 1: /cgn2_6/ptodata/2/pna/PCT_NEM_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US11_NEW_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US12_NEW_COMB.seq.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	62.8	19.4	591	6 US-10-106-698-858
2	56.8	17.6	526	6 US-10-106-698-2440
3	44.6	13.8	1186	6 US-10-103-313-621
4	44.6	13.8	1186	6 US-10-103-313-621
5	38.6	12.0	1327	6 US-10-103-313-621
6	37.1	11.5	1182	6 US-10-103-313-621
7	37.1	11.5	1182	6 US-10-103-313-621
8	31.4	9.7	988	6 US-10-106-698-892
9	30.6	9.5	254	5 US-09-975-254-16019
10	30.2	9.3	35959	6 US-10-105-299-10167
11	29.4	9.1	589	6 US-10-015-219-101
12	29.4	9.1	1103	6 US-10-105-299-2978
13	29.4	9.1	4687	6 US-10-105-299-14978
14	29.4	9.1	1694	1 PCT-US02-05912-8
15	29.4	9.0	1465	5 US-09-975-254-26080
16	29.4	9.0	1465	5 US-09-975-254-26080
17	29.4	9.0	37187	6 US-10-102-627-109
18	28.8	8.9	231	5 US-09-975-254-23269
19	28.8	8.9	1022	6 US-10-105-299-1724
20	28.8	8.9	14036	6 US-10-105-299-9008
21	28.8	8.9	14036	6 US-10-105-299-10472
22	28.8	8.9	14036	6 US-10-105-299-10518
23	28.6	8.9	257	5 US-09-975-254-5489
24	28.2	8.7	31169	6 US-10-105-299-9006
25	28.2	8.7	31169	6 US-10-105-299-9006
26	28.2	8.7	31169	6 US-10-105-299-9006
27	28.2	8.7	31169	6 US-10-105-299-9006
28	28.2	8.7	31169	6 US-10-105-299-9006
29	28.2	8.7	31169	6 US-10-105-299-9006
30	28.2	8.7	31169	6 US-10-105-299-9006
31	28.2	8.7	31169	6 US-10-105-299-9006
32	28.2	8.7	31169	6 US-10-105-299-9006
33	28.2	8.7	31169	6 US-10-105-299-9006
34	28.2	8.7	31169	6 US-10-105-299-9006
35	28.2	8.7	31169	6 US-10-105-299-9006
36	28.2	8.7	31169	6 US-10-105-299-9006
37	28.2	8.7	31169	6 US-10-105-299-9006
38	28.2	8.7	31169	6 US-10-105-299-9006
39	28.2	8.7	31169	6 US-10-105-299-9006
40	28.2	8.7	31169	6 US-10-105-299-9006
41	28.2	8.7	31169	6 US-10-105-299-9006
42	28.2	8.7	31169	6 US-10-105-299-9006
43	28.2	8.7	31169	6 US-10-105-299-9006
44	28.2	8.7	31169	6 US-10-105-299-9006
45	28.2	8.7	31169	6 US-10-105-299-9006

Sequence 12476, A
Sequence 4, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 1234, A
Sequence 15095, A
Sequence 3035, Ap
Sequence 920, Ap
Sequence 120, Ap
Sequence 2163, Ap
Sequence 14218, A
Sequence 14221, A
Sequence 14224, A
Sequence 282, App
Sequence 907, App
Sequence 149, Ap
Sequence 14155, A

ALIGNMENTS

RESULT 1
US-10-105-698-858
: Sequence 858, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al
: TITLE: Polynucleotide and Colon Cancer Associated Polynucleotides and Polypept
: PCT NUMBER: PCT/US00/26524
: CURRENT APPLICATION NUMBER: US/10106698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 856
: SOURCE: Genbank Ver. 3.0
: SEQ ID NO 858
: LENGTH: 591
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: (132)..(132)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (542)..(542)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (550)..(550)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (579)..(579)
: OTHER INFORMATION: n equals a.t.g. or c
: NAME/KEY: misc.feature
: LOCATION: (590)..(590)
: OTHER INFORMATION: n equals a.t.g. or c
US-10-106-698-858

Query Match 19.4% Score 62.8; DB 6; Length 591;
Best Local Similarity 58.3%; Pred. No. 3.1e-10;
Matches 109; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
Oy 75 gtccagagagattgaacgtatgattgagagaaaccgagccctgcatgacagcagg 134
Db 7 gtccacagagatcgacgacgtatgattgagcagggtgcgcgacccagagatggaccag 66
Oy 135 agccatgcctcacacagatctgtgtgacacagggtccagagatgaccttgccttctc 194

;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: P1207C1
;; CURRENT APPLICATION NUMBER: US/10/103,313
;; CURRENT FILING DATE: 2002-03-12
;; NUMBER OF SEQ ID NOS: 653
;; SOFTWARE: PatentIn Ver. 2.0
;; Prior Application removed - See File Wrapper or Palm
;; SEQ ID NO 622
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-103-313-622

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
Oy 80 ggaagagattgaacgtgtgattgagcagaacccgagccctgcatgcaagacagagagcca 139
Db 1182 GCAGGAGCTGGACCTGTGTGGTAGGGTGGAGGCCCGCCCAAGCCTGGACTATCGCGTGTG 1123
Oy 140 catgacctacacagatgctgtgtagcagaggtccagagataccttgacctttctccccc 199
Db 1122 CTGCGCTACCGAAGCGAGTGTCTGGAGTCCAGTGTCTCAACAGTGTGCGCTT 1063
Oy 200 cagctgcccctacagatgacctgagcatttaaatcagaactatctattcccaagg 258
Db 1062 GGGGCTGGCGCGACCTCCACCTCGACACCCAGCCTGACAGCCACTGTCTGCCCAAG 1004

RESULT 5
US-10-103-313-7201/c
;; Sequence 7201, Application US/10105299
;; GENERAL INFORMATION: st. al
;; TITLE OF INVENTION: Human Secreted Proteins
;; FILE REFERENCE: PS950
;; CURRENT APPLICATION NUMBER: US/10/105,299
;; CURRENT FILING DATE: 2002-03-26
;; NUMBER OF SEQ ID NOS: 15197
;; SOFTWARE: PatentIn Ver. 2.0
;; Prior Application removed - See File Wrapper or Palm
;; SEQ ID NO 7201
;; LENGTH: 13327
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-105-299-7201

Query Match 12.0% Score 38.6; DB 6; Length 13327;
Best Local Similarity 50.8%; Pred. No. 0.028;
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
Oy 80 ggaagagattgaacgtgtgattgagcagaacccgagccctgcatgcaagacagagagcca 139
Db 9351 GCAGGAGCTGGACCTGTGTGGTAGGGTGGAGGCCCGCCCAAGCCTGGACTATCGCGTGTG 9292
Oy 140 catgacctacacagatgctgtgtagcagaggtccagagataccttgacctttctccccc 199
Db 9291 CTGCGCTACCGCAAGCAGTGTCTGGAGATCCAGTGTCTCAACAGCGTGTGCGCTT 9232
Oy 200 cagctgcccctacagatgacctgagcatttaaatcagaactatctattcccaagg 259
Db 9231 GGGTGGCGCGACCTTCACCTCGACACCCACTGTGCACAGCCACTGTCTGCCCAAGGT 9172
Oy 260 a 260
Db 9171 A 9171

RESULT 6
US-10-103-313-622
;; Sequence 622, Application US/10103313
;; GENERAL INFORMATION: st. al
;; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;; FILE REFERENCE: P1207C1
;; CURRENT APPLICATION NUMBER: US/10/103,313
;; CURRENT FILING DATE: 2002-03-12
;; NUMBER OF SEQ ID NOS: 653
;; SOFTWARE: PatentIn Ver. 2.0
;; Prior Application removed - See File Wrapper or Palm
;; SEQ ID NO 622
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-103-313-622

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
Oy 80 ggaagagattgaacgtgtgattgagcagaacccgagccctgcatgcaagacagagagcca 139
Db 5 ggaagagattgaacgtgtgattgagcagaacccgagccctgcatgcaagacagagagcca 64
Oy 140 catgacctacacagatgctgtgtagcagaggtccagagataccttgacctttctccccc 199
Db 65 cctgacctacagcagatgctgtgtagcagaggtccagagataccttgacctttctccccc 124
Oy 200 cagctgcccctacagatgacctgagcatttaaatcagaactatctattcccaagg 259
Db 125 ggggtgctgagcagccttccaccttgcagacacccacctgcaagccactgtctgcccaagg 184
Oy 260 a 260
Db 185 a 185

RESULT 7
US-10-105-299-7203/c
;; Sequence 7203, Application US/10105299
;; GENERAL INFORMATION: st. al
;; TITLE OF INVENTION: Human Secreted Proteins
;; FILE REFERENCE: PS950
;; CURRENT APPLICATION NUMBER: US/10/105,299
;; CURRENT FILING DATE: 2002-03-26
;; NUMBER OF SEQ ID NOS: 15197
;; SOFTWARE: PatentIn Ver. 2.0
;; Prior Application removed - See File Wrapper or Palm
;; SEQ ID NO 7203
;; LENGTH: 1182
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-105-299-7203

Query Match 11.5% Score 37; DB 6; Length 1182;
Best Local Similarity 50.3%; Pred. No. 0.03; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
Oy 80 ggaagagattgaacgtgtgattgagcagaacccgagccctgcatgcaagacagagagcca 139
Db 1178 GCAGGAGCTGGACCTGTGTGGTAGGGTGGAGGCCCGCCCAAGCCTGGACTATCGCGTGTG 1119
Oy 140 catgacctacacagatgctgtgtagcagaggtccagagataccttgacctttctccccc 199
Db 1118 CTGCGCTACCGCAAGCAGTGTCTGGAGATCCAGTGTCTCAACAGCGTGTGCGCTT 1059
Oy 200 cagctgcccctacagatgacctgagcatttaaatcagaactatctattcccaagg 259
Db 1058 GGGTGGCGCGACCTTCACCTCGACACCCACTGTGCACAGCCACTGTCTGCCCAAGGT 999
Oy 260 a 260
Db 998 A 998

[illegible]

Search completed: April 19, 2002, 11:07:06
Job time: 10746 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: April 19, 2002, 07:21:40 ; Search time 4168.4 Seconds
(without alignments)
6910.103 Million cell updates/sec

Title US-09-763-292-1
Sequence: 1 cttcaatgagctctttgtg.....acatattattataaataaga 1746

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248599755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:
1: gb_ba.*
2: gb_hg.*
3: gb_mv.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl.*
8: gb_pr.*
9: gb_ro.*
10: gb_ro.*
11: gb_ro.*
12: gb_ro.*
13: gb_ro.*
14: gb_ro.*
15: gb_ro.*
16: gb_ro.*
17: gb_ro.*
18: gb_ro.*
19: gb_ro.*
20: gb_ro.*
21: gb_ro.*
22: gb_ro.*
23: gb_ro.*
24: gb_ro.*
25: gb_ro.*
26: gb_ro.*
27: gb_ro.*
28: gb_ro.*
29: gb_ro.*
30: gb_ro.*
31: gb_ro.*
32: gb_ro.*
33: gb_ro.*
34: gb_ro.*
35: gb_ro.*
36: gb_ro.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1739.5	99.5	1746	6	AR071575	AR071575 Sequence
2	1739.5	99.5	1746	6	HUMCYPC219	M61854 Human cytoc
3	1532.4	93.5	1669	6	E14931	E14931 Human mRNA
4	1534.5	97.9	1845	6	HUMCYPC29A	M61857 Human cytoc
5	1534.5	97.9	1845	6	AR071579	AR071579 Sequence
6	1529.8	97.5	1854	6	AR071576	AR071576 Sequence
7	1529.8	97.5	1854	6	AR071576	AR071576 Sequence
8	1509.8	96.3	1843	6	E022719	M61858 Human cytoc
9	1509.8	96.3	1843	6	E022719	M61858 Human cytoc
10	1495.8	95.7	1814	9	S46963	P00173 Homo sapien
11	1458.5	91.5	1473	22	E10866	S46963 putative CY
12	1456.4	91.4	2395	6	HUMSVF450A	E10866 cDNA encodi
13	1436.4	82.3	1892	6	AR071581	L07093 Human cytoc
14	1413.6	81.0	1444	6	E14930	AR071581 Sequence
15	1367.6	78.3	1577	9	HUMCYPMPA	E14930 Artificial
16	1366.6	78.3	1576	9	HUMCYPMPP	M21939 Human cytoc
17	1343.4	76.9	1473	22	E10867	M15331 Human liver
18	1343.4	76.9	1473	22	E10867	E10867 Human cDNA
19	1264.5	72.4	2009	6	HUMCYPC218	E10863 cDNA encodi
20	1264.5	72.4	2009	6	AR071577	M61856 Human cytoc
21	1264.5	72.4	2258	6	AR071580	AR071577 Sequence
22	1238.8	71.0	1346	9	HUMCYPC18	AR071580 Sequence
23	1238.8	71.0	1346	9	HUMCYPC18	M61853 Human cytoc
24	1183.2	57.8	1276	9	HUMCYPC217	X65962 H.sapiens m
25	1175.6	57.3	1866	9	HSJIC2	M61858 Human cytoc
26	1174.5	57.2	1901	9	SS3046	Y00498 H. sapiens
27	1173.8	57.2	1473	22	E10867	SS3046 cytochrome
28	1173.8	57.2	1473	22	E10867	E10867 cDNA encodi
29	1170.0	57.0	1829	9	HUMCYPAK	M61857 Human cytoc
30	1142.2	65.4	1807	9	HUMCYPMPC	M1797 Human cytoc
31	1138.6	65.2	1441	9	HUMCYPMFB	M21941 Human cytoc
32	1121.6	64.2	1868	9	HUMCYPB	M21940 Human cytoc
33	1109.8	63.6	1473	22	E10639	M17398 Human cytoc
34	1109.8	63.6	1473	22	E10862	E10639 Human cDNA
35	1106.6	63.4	1473	22	E10863	E10862 cDNA encodi
36	1105.5	63.3	1473	22	E10864	E10863 cDNA encodi
37	1097.8	62.9	1737	9	HUMCYPMFD	E10864 cDNA encodi
38	1076.4	61.6	1737	9	PABCY205	M21942 Human cytoc
39	1076.4	61.6	1737	9	PABCY205	M21942 Human cytoc
40	1065.1	61.0	1662	4	PABCY4503	M55664 Rabbit cyto
41	1051.2	60.8	1827	4	PABCY4503	D00190 Orcytolagus
42	1055.2	60.4	1871	4	PABCY4501	J02716 Rabbit cyto
43	1050.0	60.1	1679	4	AF016248	M11299 Rabbit live
44	1038.6	59.5	1758	4	SSC15D09	AF016248 Canis fam
45	1038.2	59.5	1808	10	HANP4502	Z93100 S.scrofa CY
						D11435 Mesocricetu

ALIGNMENTS

RESULT 1					
AR071575	AP071575	1746 bp	DNA	PAT	18-FEB-2000
LOCUS	Sequence 2 from patent US 5912120.				
DEFINITION	AR071575				
ACCESSION	AR071575				
VERSION	AP071575.1	GI:7222463			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unpublished.				
AUTHORS	Goldstein, J. A. and De Morais, S.M.F.				
TITLE	Cloning, expression and diagnosis of human cytochrome P450 2C19: the principal determinant of s-mephenytoin metabolism				
JOURNAL	Patent: US 5912120-A 2 15-JUN-1999;				
FEATURES	Location/Qualifiers				
source	1..1746				
BASE COUNT	477 a 418 c 365 g 486 t				
ORIGIN	/organism="unknown"				

[illegible]

Oy	1021	Tgaacagagaggagccatggcctctacacagatgtgtgagcagaggtccagagatac	1080			
Db	1021	TGCAGGACAGGGGCCACATGCCCTACAGATGCTGTGTCGACAGGCTCCAGATACA	1080			
Oy	1081	Lgacatctatcccacacacagcttcccacagtgagctgtagcttaaatlcagaact	1140			
Db	1081	TGAGACTATCCCAACAGCTGCCCATGAGTACGCTGTGACGTTAAATTCGAAACT	1140			
Oy	1141	acctatctccacacacacatataaacttccctctcccttgcctgactcagacca	1200			
Db	1141	ACCTATCTCCAGGGGCACACCATATTACTTCCTCTATCTGTGCTACATGACACA	1200			
Oy	1201	agaagatttcccacacccagagatgttaaccctcgaacttctcgatgagagtggaatt	1260			
Db	1201	AAGAAATTCCCAACACAGATGTTTGACCTCGTCACTTTCTGATGAAGGTGGAAAT	1260			
Oy	1261	ttaagaaagtaactacttcctgctcttctcagcaggaacacagattgtgtggagag	1320			
Db	1261	TTAAGAAAGTAACACTCTCACTGCTTTCTACAGAGAAACGATTTGTGTGGAGGG	1320			
Oy	1321	gcttgagccctatagagctggtttattctcagacacttcattttaaagacacttaac	1380			
Db	1321	GCCTGAGCCCTATAGAGCTGTTTATTCTCGACCTTATTACAGAACTTTAACCTGA	1380			
Oy	1381	aattctatgtatgaccacaaagacacttgacacactctctgttgaatgattgctctg	1440			
Db	1381	AAATCTGATGATGACCAAGAGACCTTGACAACTCTGTGTCATGATGATGCTGTG	1440			
Oy	1441	tcscgctctctcagctgagcttccctctcctcagagacagagctgctgcagtcg	1500			
Db	1441	TCGCCGCCCTCTATACGCTGTGCTCATCTCTGTCGAAGAAGCACAGATGGCTGCTG	1500			
Oy	1501	ctctgtgctgctctgagctcttctctctgtgacaaattcactactctataagttac	1560			
Db	1501	CTCTGTGCTGCTCAGAGCTCTTCTCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG	1560			
Oy	1561	lettctgagcctctatctcaattctctctccacagagctcagtgacactcagcttc	1620			
Db	1561	TTCTCTGACCCGCTGATCCACATTTCCCTCTCCCCAGACTAGTGAACATTCAGCTC	1620			
Oy	1621	catlaaaaggtttcaetgacaaatatactgtctatcccacactcactataagttac	1680			
Db	1621	CATTAAGAAAGTTCACTGCTGCAATATCTCTGATCTTCCCTACTCATTAATAGTCT	1680			
Oy	1681	atgttgacacacatctatcactctcaattgactgaattgacttaaacattattat	1740			
Db	1681	ATTGAGTCCCAATATGCTGATAGTCTCTAAATGTTGAGTTATTAACATATTATTATTA	1740			
Oy	1741	aataga 1746				
Db	1741					
Db	1741					
RESULT 2						
HUMCYPC219						
LOCUS						
DEFINITION	HUMCYPC219 1746 bp mRNA					
ACCESSION	Human cytochrome P4502C19 (CYP2C19) mRNA, clone 11a.					
VERSION	M61854.1 J05326					
KEYWORDS	M61854.1 GI:181343					
FEATURES	cytochrome P450; cytochrome P450 2C19.					
SOURCE	human, GONA					
ORGANISM	Homo sapiens					
REFERENCE						
AUTHORS						
TITLE						
Cloning and expression of complementary DNAs for multiple members						
of the human cytochrome P4502C subfamily [published erratum						
in Biochemistry 30 (13), 3247-3255 (1991)]						
JOURNAL						
MEDLINE						
91182240						
REFERENCE						
AUTHORS						
TITLE						
Cloning and expression of complementary DNAs for multiple members						
of the human cytochrome P4502C subfamily [published erratum						
in Biochemistry 30 (13), 3247-3255 (1991)]						
JOURNAL						
MEDLINE						
91182240						
REFERENCE						
AUTHORS						
TITLE						
Cloning and expression of complementary DNAs for multiple members						
of the human cytochrome P4502C subfamily [published erratum						
in Biochemistry 30 (13), 3247-3255 (1991)]						
JOURNAL						
MEDLINE						
91182240						
REFERENCE						
AUTHORS						
TITLE						
Cloning and expression of complementary DNAs for multiple members						
of the human cytochrome P4502C subfamily [published erratum						
in Biochemistry 30 (13), 3247-3255 (1991)]						
JOURNAL						
MEDLINE						
91182240						

RESULT	2
HUMCYPC219	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	


```

Db 1321 GCTGCGCCGATCGAGCTGTTTATTCCTGACCTTCATTTACAGCACTTTAACTCGA 1380
Oy 1381 aactcttgatgacccaagaagaccttaacaaactccctgtgtcaatgattgctctg 1440
Db 1381 AATCTCTGATTCACCAAGAGACCTTGACCAACTCTGTGTGTCATGGAATGCTCTG 1440
Oy 1441 tccgcctcttatccgtgctgtcttccctctctctctctctctctctctctctctct 1500
Db 1441 TCCGCCCTCTCTACGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1500
Oy 1501 ctccgtgctgctccgtgctgctctctctctctctctctctctctctctctctct 1560
Db 1501 CTCTGCTGCTGCTCCCTGAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1560
Oy 1561 tctctgacccgtcatctcatctctctctctctctctctctctctctctctctctct 1620
Db 1561 TCTCTGACCGCTCATCTCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
Oy 1621 catctgagatgctccgtgctgctgctgctgctgctgctgctgctgctgctgctg 1680
Db 1621 CATTAAAAAGTTTCACTGTCGAATATCTGCTATCTCCCATACATCTATATAGTAC 1680
Oy 1681 attgagtgccacaaatgctgctgctgctgctgctgctgctgctgctgctgctg 1740
Db 1681 ATTGAGTGCCCAATATGCTGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Oy 1741 aataga 1746
Db 1741 AATAGA 1746

RESULT 3
LOCUS E14931 1669 bp DNA PAT 28-JUL-1999
DEFINITION Human mRNA for hepatic cytochrome P450 2C19.
ACCESSION E14931
VERSION 1 GI:5709614
XREF NCBI GI:5709614
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Baba,T., Kirita,S. and Aoyama,J.
TITLE MASS EXPRESSION SYSTEM OF MODIFIED SUBSTANCE OF CYTOCHROME P450 2C19
IN ESCHERICHIA COLI
JOURNAL Parent: JP 1998033166-A 2 10-FEB-1998;
COMMENT SRIWONGI & CO LTD
DB JPN 1998033166-A/2
PK 10-FEB-1998
PF 23-JUL-1996 JP 1996193015
PI BABA TAKAHIKO, KIRITA SHIRO, AOYAMA JUNKO
PC C12N9/02,C07H1/04,C07K1/16,C12N1/21,C12N15/09,C12N5/02, PC
C12H1/19,
PC (C12N1/21,C12H1/19);
CC strandedness: Double;
CO topology: Linear;
Key Location/Qualifiers
FT source 1..1669
FT /organism="Homo sapiens"
FT /tissue_type="liver"
FT CDS 1..1473
FT /product="hepatic cytochrome P450 2C19"
FT Location/Qualifiers
FT 1669
FT /organism="Homo sapiens"
FT /db_xref="taxon:9606"
BASE COUNT 449 a 405 c 363 g 452 t
ORIGIN

```

```

Query Match 93.5% Score 1632.4 DB 6 Length 1669
Best Local Similarity 99.5% Pred. No. 0
Matches 1636: Conservative 0 Mismatches 6 Indels 0 Gaps 0
Oy 1 ctccaatgagatccctttatgagctctgtgtctgtctctctctctctctctctctca 60
Db 2 CTTCATGATGCTTTTGTGTCCTTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 87
Oy 61 tctgagagacagactctgagagagagaaactccctctgagccactctctctccatga 120
Db 61 TCTGAGACAGAGCTCTGGGAGAGAAAACCTCTCTCTCTCTCTCTCTCTCTCTCTCT 147
Oy 121 ttggaatatctacagatagatatagagatgctcagcaaatctcttaaccaatctctca 180
Db 148 TTGGAATACTCTACAGATAGATATAGAGATGTCAGCAAACTCTTACCACTCTCTCA 207
Oy 181 aaactatagctctgtctctctctctctctctctctctctctctctctctctctct 240
Db 208 AATCTATGCGCTGTCTCTCTCTGTATTTTGGCTTGGAGCATGCTGTCTCTCTCT 267
Oy 241 gatatgagtggtgagagagagccctctctctctctctctctctctctctctctct 300
Db 268 GATATGAAGTGGTGAAGAGAGCCCTGATTTGATCTTGGAGAGAGCTTTCTGGAAGGCC 327
Oy 301 attctccactggctgagagagctctcaacagagattggactctgttctcaagatggaaga 360
Db 328 ATTCTCCACTGGCTGAAAGAGCTTACAGAGGTTTGGAAATCTTTTACAGATGGAGAA 387
Oy 361 gatgagagagattcggcctttctctctctctctctctctctctctctctctctct 420
Db 388 GATGCAAGGAGATCCGGCTTTCTCTCTCATGAGCTCGGGAATTTGGGATGGGGAAGA 447
Oy 421 gtagcatgagagacctctctcaagagagagccctgtctgtgagagagttgagaaaa 480
Db 448 GAGCATGTGAGACCTGTCTCAAGAGAGAGCCCTCTCTCTGTGGAGAGTTTGAAGAAA 507
Oy 481 ccaagctctccctgctgctctctctctctctctctctctctctctctctctctctct 540
Db 508 CCAAGCTTCACTGTGATCCCACTTTCTCTCTGCTGCTGCTCTCTCTCTCTCTCTCT 567
Oy 541 gctccattatttccagaaactgttctctctctctctctctctctctctctctctct 600
Db 568 GCTCCATTATTTCCAGAAAGCTTTGATTTAAGATCAGCAATTTCTTAACTTGATGG 627
Oy 601 aaaaatgagatgaacactcaggatgtagaacacccctctgagatcagatagcaaatl 660
Db 628 AAAAATGAATGAACATCAGATTTGAGACCCCTTGGATCCAGATGCAATTAAT 687
Oy 661 tttccactcattgattatttcccgagagccatacaaaactctctctctctctctct 720
Db 688 TTCCCACTATCATTTGATTTTCCCGGAACTTCACTGCTGCTGCTCTCTCTCTCTCT 747
Oy 721 ttatggaagtgatatttggagaaagttaaaagaccccaagaatcagtgagcatcaaca 780
Db 748 TTATGGAAGTGATATTTGGAGAAAGTAAAGAACACCAAGATCTGTGACATCAACA 807
Oy 781 acctctggagactttatctctctctctctctctctctctctctctctctctctct 840
Db 808 ACCCTCGGAGCTTTATGATTTCTCTCTGATCAAAATGGAGAGAAAGCAAAACCAAC 867
Oy 841 agtctgaatcactattgaaactctgtactctgactgacttacttctgagctggga 900
Db 868 AGTCTGAATTCATTTGAAAGCTTTGGTAATCTCTGAGCTGACTTACTTGGACTGGGA 927
Oy 901 cagagacacagacacactctgagatagctctctctctctctctctctctctctctct 960
Db 928 CAGACACACAGCAACCTCTGAGATAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 987
Oy 961 tcacagataagctccagagagagattgaacagtgctctctctctctctctctctctct 1020
Db 988 TCACAGTAAAGTCCAGGAGATGTAACGTTGCTGTGTGGAGAAAGAGCCGCTCTCA 1047

```


hydroxylase activity
 Biochem. Int. 27 (5), 1073-1081 (1992)
 93075249
 Genbank staff at the National Library of Medicine created this
 entry [NCBI gi38911719] from the original journal article.
 The authors also give sequence of another cytochrome P-450 that
 contains a 6-base deletion of nucleotides 783-788.
 FEATURES
 Location/Qualifiers
 1..1814
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..1814
 /gene="putative CYP2C9 gene"
 3..1436
 /gene="putative CYP2C9 gene"
 /codon_start=-1
 /product="cytochrome P-450"
 /protein_id="BAB23864.2"
 /db_xref="GI:5705937"
 /translation="MILLSLRSSGRKLPKPGTPLPVIGNILQIKDISKSLNL
 SKYGVETLYGLKPIVYVHLGTEAVKEALIDLAEFGSICRGIPLAEPANRFGIVFS
 NGKKKEIRFSLMTLNFGKRSIEDRYOEACLVLEELIKRASPDIYDFLUGCA
 PLKVAIKSIFRREDYKDOOFLNMEKLNKILSLSPMIOKNNFSPIDYDFGTN
 KLLANFKMSIDERYKESDMNNPDYDFLPMKEXKRNQSPFIESLEN
 TDAVVRVQVDTSTYRALLKLLHPYATVQSEIIEVIRVGRNPSCHDRSIMPY
 MFDPIHFLDSEGNFKSKYFSPSAGRIICVGEALAGNELLEFLISLONPLKSLVD
 PENLDTTPVYNGFASVPPYDGLFIPV"
 BASE COUNT 517 a 413 c 369 g 515 t
 ORIGIN
 Query Match 85.7% Score 1495.8; DB 9; Length 1814;
 Best Local Similarity 92.8%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 107; Indels 17; Gaps 2;
 Oy 43 gttctctctcttcaatcgtgacacagcctctgagagagaaactccctcagcc 102
 Db 1 GTTTGCTCTCTCTTCACTCTGGACACAGAGCTCTGGAGAGGAAACTCCTCTCGCC 60
 Oy 103 ccaactctctccagtgatggaatactctacagatagatttaaggatgctcagcaat 162
 Db 61 CCACCTCTCTCCAGGATGGAAATATCTACAGATAGGTATTAGGACATCAGCAAT 120
 Oy 163 ccttaacaaactctcaaaaactctatgacctgttcaactctgatttggcctgaac 222
 Db 121 CTTTAAACCAATCTCTCAAGGCTCTAAGCCCTGTCTCACTCTCTTTTGGCTTAAC 180
 Oy 223 gcatgggtgctgcatgatatgaagtgatgaggaagcctcctgatctctggagag 282
 Db 181 CCATAGTGTGCTGCATGGATATGAACAGTGAAGGAGCCCTGATTGATCTTGGAGG 240
 Oy 283 agtttctggagagggccatttcccactggctgaagagcgttaacagagagatttggaa 342
 Db 241 AGTTTCTGGAGAGGCAATTTCCCACTGGCTGAAGAGCTTACAGAGATTTGGAA 300
 Oy 343 ttttcgaagtgaggaagatggaagagattcgcgtttctccctcactgacagtcaga 402
 Db 301 TTTTCAGCAATGGAAGAAATGGAAGGAGATCCGGGCTTTCTCCCTCATGAGCTCG 360
 Oy 403 atttggagtgaggaagagcattgagagcgtgttcaagaggaagcctgcgtccttg 462
 Db 361 ATTTTGGGATGGGGAAGAGGAGGATGAGGACCGGTTCAGAGGAGAGCCCGCTG 420
 Oy 463 tggagagtgaggaagaaacaaagcttcaacctgtgacccacttctcctggcgtg 522
 Db 421 TGGAGAGTTTGAAGAAACCAAGGCTTCACTCTGTATCCCACTTTTCACTTGGGCTG 480
 Oy 523 ctccctcgaatgctcgtctcatttctccagaaacgtttcgattataaatacag 582
 Db 481 CTCCCTGCAATGATCTGCTCCATTATTTCCATAACGTTTTCATTATTAAGTACAG 540

Oy 583 aatttcttaacttgatgaaaaaattgaatgaaaaacatcagatgtgaagcaccctcaga 642
 Db 541 AATTTCTTAACCTTAATGAAAGATTGAATAAACAATCAAGATTTTTCAGCAGCCCTGGA 600
 Oy 643 tccagatagcaataatttttccctcactcattgatttatttccctggggaaccataaacaat 702
 Db 600 TCCAGTCTGCAATATTTTCTCATCATGATTACTTCCGGGAGATCACAACAAT 660
 Oy 703 tacttaaaaactcgtctttatggaagatgattttgggaaaaagtaaaaagacaccaa 762
 Db 660 TCCAGTCTGCAATATTTTCTCATCATGATTACTTCCGGGAGATCACAACAAT 660
 Oy 762 ttt 822
 Db 661 TACTTAAAACGTTGCTTTTATGAAAAGTTATTTTGGAAAAGTAAAGAACCACCAAG 720
 Oy 763 aatcgtatgacatcaacaacccctcggaactttatgattgcttctctgatacaaatggaga 822
 Db 721 AATCAATGGACATGAACAACCCCTCAGGACTTTATTTGATCTCTCTGATGAATGGAGA 780
 Oy 823 aggaagacaaacaaacacagctgattcactatgaaaacttggtaactcactgcagctg 882
 Db 781 AGGAAGAAGCACACACACATCTGAATTTACTATTGAAAGCTTGGAAACACTGCAAGTT 840
 Oy 883 actactctggagctgggagacacacacacacacacacacacacacacacacacacac 942
 Db 841 ACTTGTGTGAGCTGGGACAGACAGACAGACAGACACACACACACCTTGAGATGCTCTCTCTCC 900
 Oy 943 tcttaagac 1002
 Db 901 TCTTGACACCCACAGAGGTACAGCTTAAAGTCCAGGAGAGATGAACGCTGTGATTGGCA 960
 Oy 1003 gaaacagagacccctctgctacacagacagagggcgcacactcctcaacagatccttga 1062
 Db 961 GAAACCGAGCCCTGCATGACAGACAGAGCCACACATGCCCTACACAGATGCTGTGTTGC 1020
 Oy 1063 acaggtccagagatacactcagacctcatcccccacagcctgcgccatgcaatgaacctatg 1122
 Db 1021 AGAGGTCCAGATACATTTGACCTTCTCCACACAGCTGCCCCATCGCATGAGCTGTG 1080
 Oy 1123 acgttaactcagaactcactcacttcccaaggacacacacacacacacacacacacacac 1182
 Db 1081 AATTAATTAAGAACTATCTATCTCCAGGGCACACACATATTAATTTCCCTGACTTT 1140
 Oy 1183 clgtgctacatgac 1242
 Db 1141 CTGTGCTACATGACACAAAGAAATTTCCCAACCCAGAGATGTTTGACCTCATCATCTTC 1200
 Oy 1243 tggatgagtggaatattttaaagaaagtaactacttcaatgcttcttccagcagagaaac 1302
 Db 1201 TGAATAGGTGGCAATTTTAAAGAAAGTAATACTTCTGCTTCTCAGCAGGAAAC 1260
 Oy 1303 gaatttctgtgagagagcctgcccagacagagctgttttcttctcactcactcactcact 1362
 Db 1261 GGAATTTGTGTGGAGAGAGCCCTGGCGGCAATGGAGCTGTTTTTATTCCTGACCTCATTT 1320
 Oy 1363 tacagaacttcaactgaaactctctgattgacccaaagagcctgacacacacacacacac 1422
 Db 1321 TACAGAACTTTAACCCTGAAATCTCTGTTGACCCAAAGAACCTTGACACACACCTCAGTTG 1380
 Oy 1423 tcaatgatttcttctgcccaccccttctcactcagctgcttctcacttctctgagaaa 1482
 Db 1381 TCATAGATTTGGCTCTGTGCGCCCTCTACAGCTGCTCTCATTTCTCTGTGAGAGA 1440
 Oy 1483 gcacagatgctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1535
 Db 1441 GACAGATGGCTGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 Oy 1536 tccaaattcactatctgatacttcttcttcttcttcttcttcttcttcttcttcttctt 1595
 Db 1501 TCAATCTTCACTATCTGATGCTTTTTCACCTGTCATCTCACAATTTTCCCTTCCTT 1560
 Oy 1596 caagctcagagaaactcagcctcacta-----aaaaagttcactcagaaa 1645
 Db 1561 GAAGATCTAGTGAACATTCGACCTCCATAGGAGAGGTTTCTCTATGTTTCACTGTGCANA 1620

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 ; Search time 277.57 Seconds
(without alignments)

Title: US-09-763-292-1

Perfect score: 1746
Sequence: 1 ctccaatgatgcctttttgtg.....acatatattattataataga 1746

Scoring table: IDENTITY_NUC

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0.9

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :	
1:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1980.DAT.*
2:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1981.DAT.*
3:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1982.DAT.*
4:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1983.DAT.*
5:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1984.DAT.*
6:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1985.DAT.*
7:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1986.DAT.*
8:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1987.DAT.*
9:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1988.DAT.*
10:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1989.DAT.*
11:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1990.DAT.*
12:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1991.DAT.*
13:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1992.DAT.*
14:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1993.DAT.*
15:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1994.DAT.*
16:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1995.DAT.*
17:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1995.DAT.*
18:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1995.DAT.*
19:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1997.DAT.*
20:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA1999.DAT.*
21:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA2000.DAT.*
22:	/\$IDS2/\$cgdata/\$geneseq/\$geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1739.6	99.6	1746	19	AAV44153	Human cytochrome P
2	1739.6	99.6	1748	22	AAK12241	Human cytochrome P
3	1738	99.5	1746	17	AAI03708	Cytochrome P450 2C
4	1632.4	87.9	1682	19	AAI22559	Wild type cytocho
5	1632.4	87.9	1682	20	AAI22559	Mammalian cytocho
6	1534.6	87.9	1582	17	AAI13765	Cytochrome P450 2C
7	1534.6	87.9	1582	17	AAI44157	Human cytochrome
8	1529.8	87.6	1654	17	AAI11378	Cytochrome P450 2C
9	1529.8	87.6	1654	17	AAV44154	Human cytochrome P
10	1458.6	83.5	1473	16	AAQ87728	Human auxiliary cy
11	1458.6	83.5	1473	17	AAI28394	Human cytochrome P

Human liver cytochrome P
Human cytochrome p
Genetic construct
Cytochrome P450 2C
Human derived cyto
Human derived cyto
Human cytochrome p
Partial sequence o
Human cytochrome p
Human cytochrome P
Human derived cyto
Human cytochrome P
Cytochrome P450 2C
Human cytochrome p
Human cytochrome p
Human auxiliary cy
Human liver cell s
Human cytochrome p
Cytochrome P450 2C
Human cytochrome p
Human auxiliary cy
Human cytochrome p
Human derived cyto
Human auxiliary cy
Human cytochrome p
Human cytochrome p
Human cytochrome p
Bacterial and mamm
Human derived cyto
Human cytochrome p
Human cytochrome p
Human derived cyto

ALIGNMENTS

RESULT 1	
AAV44153	AAV44153
ID	AAV44153 standard; CDNA; 1746 BP.
XX	
XX	AAV44153;
AC	
XX	
DT	06-OCT-1998 (first entry)
XX	
DE	Human Cytochrome P450 2C19 clone 11a CDNA.
XX	
XX	Cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
XX	Identification; mutagenic; carcinogenic; cytotoxic; haemoprotein;
XX	xenobiotic; environmental pollutant; ss.
XX	
XX	
US	Homo sapiens.
XX	
PN	US5786191-A.
XX	
PD	28-JUL-1998.
XX	
XX	
PF	22-FEB-1994; 94US-0201118.
XX	
XX	
PR	22-FEB-1994; 94US-0201118.
PR	09-APR-1992; 92US-0864962.
XX	
XX	
PA	(GOLD/) GOLDSTEIN J A.
PA	(ROMK/) ROMKES-SPARKS M.
XX	
XX	
PI	Goldstein JA, Romkes-sparks M;
XX	
XX	WPT: 1998-436528/37.
XX	
XX	
PT	Screening for drugs metabolised by cytochrome P450 - for identify
PT	mutagenic, carcinogenic or cytotoxic compounds


```

Db 486 ccaaggctccctctggtatccacatttccacatttcccttccctggaagatctagtgacatt 1625
Oy 1614 cagctccattt -----aaagtttccagtgcaaatatctgtctattcccca 1663
Db 1626 cgaacctccattcaggaagttctctatgttccacttgcaaatatctgtctattcccca 1685
Oy 1664 tactctataatgtacattgagtcacataatgtctgatactgtctatgttgagttta 1723
Db 1686 tactctgtaacagtgatgactgtccataatgtctacattctattctaatgttgagttta 1745
Oy 1724 ttaacatattattattataataga 1746
Db 1746 ttaatagttattattataataga 1768

RESULT 6
AAT11381
ID AAT11381 standard; cDNA: 1852 bp.
AC AAT11381;
XX 09-SEP-1996 (first entry)
XX Cytochrome P450 2C9 clone 65 coding sequence.
XX Cytochrome P450 2C19; human: liver; PCR: primer: detection: CYP2C19;
XX stereospecific S-mephenytoin 4'-hydroxylase activity; polymorphism: ss.
XX Homo sapiens.
XX Key Location/Qualifiers
XX CDS 11..1483
XX /tag= a
XX /product= Cytochrome P450 C9 clone 25
XX /tag= b
XX misc_difference 1085
XX /note= "Variable position"
XX /tag= c
XX /note= "Variable position"
XX W09530766-A1.
XX 16-NOV-1995.
XX 08-MAY-1995; 95WO-US05744.
XX 06-MAY-1994; 94US-0238821.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX De Morais SMF, Goldstein JA, Pomkes-Sparks M;
XX WPI: 1996-077257/08.
XX P-PSDB: AAR89865.
XX New isolated cytochrome P450 2C subfamily member - used for
XX identifying drugs metabolised by S-mephenytoin 4'-hydroxylase
XX activity and to develop other screening assays
XX
XX Example 2; Page 104; 169pp; English.
CC The sequences given in AAT11378-81 encode allelic variants of cytochrome
CC P450 2C9. The majority of clones isolated from liver S33 coded for 2C9.
CC Of the 50 clones encoding 2C9, only two allelic variants were found. 39
CC of the 2C9 clones were identical with clone 65, and 11 were identical
CC with clone 25. Clones 25 and 65 are identical in the 5' and
CC 3'-noncoding regions but contained 2 single base changes at positions
CC 1075 and 1425. One of these base changes was conservative but the other
CC results in one amino acid difference at position 359 (Ile to Leu). The
CC clones 25 and 65 refer to one nucleotide difference at position 1154
CC (G to A) which results in a single amino acid change, Thr to Met
CC at position 385. Clone 29c has a very long, 198 bp, 5'-noncoding region

```


XX Cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
KW identification; mutagenic; carcinogenic; cytotoxic; hemoprotein;
XX xenobiotic; environmental pollutant; ss.
XX Homo sapiens.
OS
XX US5786191-A.
XX 28-JUL-1998.
XX 22-FEB-1994; 94US-0201118.
XX 22-FEB-1994; 94US-0201118.
XX 09-APR-1992; 92US-0864962.
XX (GOLD)/ GOLDSTEIN J A.
XX (ROMK)/ ROMKES-SPARKS M.
XX Goldstein JA, Romkes-sparks M;
PI WPI; 1998-436528/37.
XX Screening for drugs metabolised by cytochrome P450 - for identifying
XX mutagenic, carcinogenic, or cytotoxic compounds
XX Example 2; Column 53-56; 63pp; English.
XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated
CC from clone 65. This polypeptide is used in a method to screen for a drug
CC that is metabolised by a cytochrome P450. The protein can also be used to identify a
CC (mutagenic, carcinogenic or cytotoxic) compound. Cytochrome P450 are a
CC large family of hemoprotein enzymes capable of metabolising xenobiotics
CC such as drugs carcinogens and environmental pollutants as well as
CC endobiotics such as steroids, fatty acids and prostaglandins.
XX Sequence 1852 BP; 514 A; 424 C; 380 G; 534 T; 0 other;
XX
Query Match 87.9%; Score 1534.6; DB 19; Length 1852;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps 2;
OY 1 cttcaatgagctcttttggttctgtctgtctctcatgtttgtctctcttcaa 60
DB 6 cttcaatgagctcttttggttctgtctgtctctcatgtttgtctctctc 65
OY 61 tctggagacagctctggagaggaagaaatccctctgagccactctccca 120
DB 66 tctggagacagctctggagaggaagaaatccctctgagccactctccca 125
OY 121 ttggaataatctcaagatagatataagagatgcagaaatctcaacatctc 180
DB 126 ttggaataatctcaagatagatataagagatgcagaaatctcaacatctc 185
OY 181 aaatctgagctctgttcaactgtatttggctggagagcagtggtgctgcatg 240
DB 186 aggtctatggtctgttcaactgtatttggctggagagcagtggtgctgcatg 245
OY 241 gatataagtggtgagagagcctgattgattcttgagagagagattttcgaagagcc 300
DB 246 gatataagtggtgagagagcctgattgattcttgagagagagattttcgaagagcc 305
OY 301 atttccactggttgaagagatcaacagagatttgaatcggtttccagcaatggaaga 360
DB 306 ttttccactggttgaagagatcaacagagatttgaatcggtttccagcaatggaaga 365
OY 361 gatgagagagattcggctttctctctctctctctctctctctctctctctctct 420
DB 366 aatggagagagatcggctttctctctctctctctctctctctctctctctctct 425
OY 421 ggaagcattgagaccgtgttcaagagagagccgctgtctgtgagagaggttgaga 480

DB 426 ggaagcattgagaccgtgttcaagagagagccgctgtctgtgagagagttgaga 485
OY 481 caagagcttcaactgtgagctctctctctctctctctctctctctctctctctct 540
DB 486 caagagcttcaactgtgagctctctctctctctctctctctctctctctctctct 545
OY 541 gttccatttatttccagaaacttttgattataaagatcaagaaatttcttaacttga 600
DB 546 gttccatttatttccagaaacttttgattataaagatcaagaaatttcttaacttga 605
OY 601 aaaaattgaatgaagaactcaagatttgaagcaccctctgagatccagatgaatatt 660
DB 606 aaaaattgaatgaagaactcaagatttgaagcaccctctgagatccagatgaatatt 665
OY 661 tttccactatatttatttctccgggaacctataaacaatttacttaaaacttgcct 720
DB 666 tttccactatatttatttctccgggaacctataaacaatttacttaaaacttgcct 725
OY 721 ttatgaaagtatatttttggaaagaaactaaagaacacacaaatgaatgaatcaaca 780
DB 726 ttatgaaagtatatttttggaaagaaactaaagaacacacaaatgaatgaatcaaca 785
OY 781 acctcggagctttatttatttctctctctcaaaatggagaagaaagaaacaaacac 840
DB 786 acctcggagctttatttatttctctctctgaagaaatggagaagaaagaaacacac 845
OY 841 agtctgaattcatttgaagaacttgaatcactgcagctgaacttatttggagctgga 900
DB 846 cacttgaatttatttgaagaacttgaatcactgcagctgaacttatttggagctgga 905
OY 901 cagagacaagaac 960
DB 906 cagagacaagaac 965
OY 961 tccagctaaagtcagagagagattgaagctgtcatttgcagaaacacagacccctcga 1020
DB 966 tccagctaaagtcagagagagattgaagctgtcatttgcagaaacacagacccctcga 1025
OY 1021 tgcacagagggccacatgccctacacagatgctgtgtgcacaggtccagagataca 1080
DB 1026 tgcacagagggccacatgccctacacagatgctgtgtgcacaggtccagagataca 1085
OY 1081 tgcacatctccacacagctgtcccaatgcagtgacctgtgaccttgaattcagaact 1140
DB 1086 tgcacatctccacacagctgtcccaatgcagtgacctgtgaccttgaattcagaact 1145
OY 1141 acctcattcccaagggacacacacacacacacacacacacacacacacacacac 1200
DB 1146 acctcattcccaagggacacacacacacacacacacacacacacacacacacac 1205
OY 1201 aagaatttcccaacacagagatgttgaacctgcacactttctgagatgaagtggaatt 1260
DB 1206 aagaatttcccaacacagagatgttgaacctgcacactttctgagatgaagtggaatt 1265
OY 1261 ttaagaagaataacttctcactcactctctcagcagaaacacacacacacacacac 1320
DB 1266 ttaagaagaataacttctcactcactctcagcagaaacacacacacacacacac 1325
OY 1321 gcttgcgcgcagagagctgttttattctgaccttcaatttcaagaactttaacctga 1380
DB 1326 ccttgcgcgcagagagctgttttattctgaccttcaatttcaagaactttaacctga 1385
OY 1381 aattctgtattgaacccaagagacacttgcacacactcctgtgttgaatgaatttgcctg 1440
DB 1386 aattctgtattgaacccaagagacacttgcacacactcctgtgttgaatgaatttgcctg 1445
OY 1441 tcccgccttctacagctgtgtgttctcctgtctgagagacacagatggttgcctg 1500
DB 1446 tcccgccttctacagctgtgtgttctcctgtctgagagacacagatggttgcctg 1505
OY 1501 cttctgtgtctctcagcctctcttctctctctg-----tccaaatttcaactatctg 1553
DB 1501 cttctgtgtctctcagcctctcttctctctctg-----tccaaatttcaactatctg 1553


```

Db 788 accctcaggacttattgattgctctcctgatgaaatggagaaagcaacacac 847
Oy 841 agtctgaattcattgtaaaacttggtaactcactgagctgacttacttggagctgga 900
De 848 cactcgaattactcattgaagctggaaacactcgtgacttcttggagctgga 907
Oy 901 cagagcaacacacacacacacacacacacacacacacacacacacacacacac 960
Oy 908 cagagcaacacacacacacacacacacacacacacacacacacacacacacac 967
Oy 961 tcacagctaaagtccagagagagatgaaacgtgctatggcagaacacacacacac 1020
Db 968 tcacagctaaagtccagagagagatgaaacgtgctatggcagaacacacacacac 1027
Oy 1021 tcacagcagaggggcccacatgccctcaccagatgctgctgacagagacacagataca 1080
Db 1028 tcagagcagagggcccacatgccctcaccagatgctgctgacagagacacagataca 1087
Oy 1081 tcagcctcattccacacacacacacacacacacacacacacacacacacacacac 1140
Db 1088 tcagcctcattccacacacacacacacacacacacacacacacacacacacacac 1147
Oy 1141 acctcatttccacacacacacacacacacacacacacacacacacacacacacac 1200
Db 1148 atctcatttccacacacacacacacacacacacacacacacacacacacacacac 1207
Oy 1201 agaatcttccacacacacacacacacacacacacacacacacacacacacacac 1250
Db 1208 agaatcttccacacacacacacacacacacacacacacacacacacacacacac 1257
Oy 1261 taaagaagaataactcattcattcctcagcagagaaacagatttctgaggagag 1320
Db 1268 taaagaagaataactcattcattcctcagcagagaaacagatttctgaggagag 1327
Oy 1321 gcttggcagatgagctgcttcttctcagcactcatttcaagaacatttaacctga 1380
Db 1328 ccttggcagatgagctgcttcttctcagcactcatttcaagaacatttaacctga 1387
Oy 1381 aactctgctgagcagcagacacacacacacacacacacacacacacacacacac 1440
Db 1388 aactctgctgagcagcagacacacacacacacacacacacacacacacacacac 1447
Oy 1441 tccgcgcttctcagcgtgcttctcctcctcagcagacacagatgctgctgctg 1500
Db 1448 tgcgcgcttctcagcgtgcttctcctcctcagcagacacagatgctgctgctg 1507
Oy 1501 ctctgtgctgctcctcagcgtgcttctcctcctcagcagacacagatgctgctg 1553
Db 1508 ctgctgtgctgctcctcagcgtgcttctcctcctcagcagacacagatgctgctg 1557
Oy 1554 tgaatgcttctcagcgtgcttctcctcctcagcagacacagatgctgctgctg 1613
Db 1561 tgaatgcttctcagcgtgcttctcctcctcagcagacacagatgctgctgctg 1627
Oy 1614 cagcgtcctccta-----aaagttcctcagcagacacacacacacacacacac 1663
Db 1628 cagcgtcctcctaagcgttctcctcagcagacacacacacacacacacacacacac 1687
Oy 1664 tactctaatagttcattgagtcgacacacacacacacacacacacacacacacacac 1723
Db 1688 tactctaatagttcattgagtcgacacacacacacacacacacacacacacacacac 1747
Oy 1724 taaacattatttataataga 1746
Db 1748 taaatatttatttataataga 1770

RESULT 9
ANV4154
Db ANV4154 standard; cDNA: 1854 BP.
IX
AC ANV4154;

```

```

XX 06-OCT-1998 (first entry)
XX Human cytochrome P450 2C9 clone 25 cDNA.
XX Cytochrome P450; drug screening; S-mephenytoin 4'-hydroxylase;
XX Identification: mutagenic; carcinogenic; cytotoxic; haemoprotein;
XX xenobiotic; environmental pollutant; ss.
XX Homo sapiens.
XX U55786191-A.
XX 28-JUL-1998.
XX 22-FEB-1994; 94US-0201118.
XX 22-FEB-1994; 94US-0201118.
XX 05-APR-1992; 92US-0864962.
XX (GOLD//) GOLDSTEIN J A.
XX (ROM//) ROMKES-SPARKS M.
XX Goldstein JA. Romkes-sparks M;
XX WPI: 1998-436528/37.
XX Screening for drugs metabolised by cytochrome P450 - for identifying
XX mutagenic, carcinogenic, or cytotoxic compounds
XX Example 2: Column 39-42: 63pp: English.
XX This sequence encodes a cytochrome P450 2C9 polypeptide isolated from
XX human clone 25. This polypeptide is used in a method to screen for a drug
XX that is metabolised by a cytochrome P450 having S-mephenytoin
XX 4'-hydroxylase activity. The protein can also be used to identify a
XX mutagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a
XX large family of haemoprotein enzymes capable of metabolising xenobiotics
XX such as drugs carcinogens and environmental pollutants as well as
XX endobiotics such as steroids, fatty acids and prostaglandins.
XX Sequence 1854 BP: 513 A; 424 C; 381 G; 536 T; 0 other;

Query Match 87.6%; Score 1529.8; DB 19; Length 1854;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1634; Conservative 0; Mismatches 112; Indels 17; Gaps 2;

Oy 1 ctccaatgagatcttttggcttggcttgccttctcctcagcagacacacacacacac 60
Db 8 ctccaatgagatcttttggcttggcttgccttctcctcagcagacacacacacacac 67
Oy 61 tctgagacagagctctggagagagagagagagagagagagagagagagagagagag 120
Db 58 tctgagacagagctctggagagagagagagagagagagagagagagagagagagag 127
Oy 121 ttggaatactcctcagacatagatattaaagagatgcagcaaatccttaaccaatctc 180
Db 128 ttggaatactcctcagacatagatattaaagagatgcagcaaatccttaaccaatctc 187
Oy 121 aaactatggccttggcttgccttgccttgccttgccttgccttgccttgccttgc 240
Db 188 aggtctatggccttgccttgccttgccttgccttgccttgccttgccttgccttgc 247
Oy 241 gatatgagcagtgagagagagagagagagagagagagagagagagagagagagag 300
Db 248 gatatgagcagtgagagagagagagagagagagagagagagagagagagagagag 307
Oy 301 atttccctcgtgctgagagagagagagagagagagagagagagagagagagagag 360
Db 308 ttctccctcgtgctgagagagagagagagagagagagagagagagagagagagag 367
Oy 351 gatagaagagagagagagagagagagagagagagagagagagagagagagagagag 420

```

```

Db 368 aatgaaagagatccggcgttctccctcctcagcgtcggaattttggagtggaaga 427
Oy 421 gagacattgagacccgtgtccaagaagaccgctgctcttggaagaggttaagaasaa 480
Db 428 gagacattgagacccgtgtccaagaagaccgctgctcttggaagaggttaagaasaa 487
Oy 481 ccaagccttccacccgtgtcaccacgttccacccgtgtcaccacgttccacccgtgtc 540
Db 488 ccaagccttccacccgtgtcaccacgttccacccgtgtcaccacgttccacccgtgtc 547
Oy 541 gttccattttccagaacacgttctgataaagatcacagcattttcctaacttgagg 600
Db 548 gttccattttccagaacacgttctgataaagatcacagcattttcctaacttgagg 607
Oy 601 aanaattgaatgaataacatcaggatttgaagccccctctggtacccagatgacataatt 660
Db 608 aanaattgaatgaataacatcaggatttgaagccccctctggtacccagatgacataatt 667
Oy 661 tccacattcattcatttcccggaacccacataaacaatttcttaaaagacttgctt 720
Db 668 tttctctcattgatttacttcccggaactcacaacaatttacttaaaagcttgctt 727
Oy 721 ttatgaaagtgatatttggagaagtaaaagaaacacccaagaatcgatgacataca 780
Db 728 ttatgaaagtgatatttggagaagtaaaagaaacacccaagaatcgatgacataca 787
Oy 781 accctcgagcttattgattgcttctcgtatcaaaatggaggaagaaagcaaaacac 840
Db 788 accctcgagcttattgattgcttctcgtatcaaaatggaggaagaaagcaaaacac 847
Oy 841 agctggaattcatttgaacacgttgatcactcagcagctgacttacttgagctggga 900
Db 848 catctgatttacttgaagcttggaacacactcagctgacttcttgagctggga 907
Oy 901 cagagacaagaacacacccctcagagacagctctctctcctcgtcagacacccagagg 960
Db 908 cagagacaagaacacacccctcagagacagctctctcctcgtcagacacccagagg 967
Oy 961 tcaagctaaagtcagagagagattgagagctgacttctgacagaaacccagccctgca 1020
Db 968 tcaagctaaagtcagagagagattgagagctgacttctgacagaaacccagccctgca 1027
Oy 1021 tgcacagacagggccacatgccctacacagatgctgtgacagaggtccagagataca 1080
Db 1028 tgcacagacagggccacatgccctacacagatgctgtgacagaggtccagagatacc 1087
Oy 1081 tgcacctcattccacacagcctgcccctgacgtgacgtgacgttcaattcagaact 1140
Db 1088 tgcacctcattccacacagcctgcccctgacgtgacgtgacgttcaattcagaact 1147
Oy 1141 acctcattcccaagggcacaacatttacttccctcacttctcgtcgtacatgacataca 1200
Db 1148 acctcattcccaagggcacaacatttacttccctcacttctcgtcgtacatgacataca 1207
Oy 1201 aagaatttcccaacccagagatgttgacctcgtcacttcttgagtgaggtggaatt 1260
Db 1208 aagaatttcccaacccagagatgttgacctcgtcacttcttgagtgaggtggaatt 1267
Oy 1261 taagaagtgtaacttacttctcgttctcagcaggaacacgattctgctggaggag 1320
Db 1268 taagaagtgtaacttacttctcgttctcagcaggaacacgattctgctggaggag 1327
Oy 1321 gctcgcgcgcatgagagctgttttttacttctgacttcttatttcaacacatttaactga 1380
Db 1328 cctcgcgcgcatgagagctgttttttacttctgacttcttatttcaacacatttaactga 1387
Oy 1381 aatctcgtatgacccaaggaacgttgacacacactcctgtgtcagatttgcctctg 1440
Db 1388 aatctcgtatgacccaaggaacgttgacacacactcctgtgtcagatttgcctctg 1447
Oy 1441 tcccgccttctcactcagctgctcacttctcgtcgtcgaagacacagatgctcgtg 1500
```

1448 tgcgccttctcaccagctgcttctccttctcgtcgtcgaagagcagctggtcgtgctg 1507
1501 cctcgtgctcctcagctccttctcctcgtg-----tccaaatttctcactatcg 1553
1508 cgtcgtgagctcctcagctccttctcctcgtggtggaattatcctccttctcactatcg 1567
1554 tgaatctctcctcagcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1613
1568 tgaatctctcctcagcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1627
1614 cagctcctcctc-----aaaaagtttctcagctggaattatcctcctcctcctc 1663
1628 cagctcctcctcagcagaggttctcctcctcctcctcctcctcctcctcctcctcctc 1687
1664 tctcctcctcctcagcagcctcctcctcctcctcctcctcctcctcctcctcctcctc 1723
1688 tctcctcctcagcagcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1747
1724 ttaacatatttatttaataaga 1746
1748 ttaatagttatttataataaga 1770

RESULT 10
AA087728
ID AA087728 standard: cDNA: 1473 BP.
AC AA087728:
XX 15-NOV-1995 (first entry)
DE Human auxiliary cytochrome P450 species 2C19 coding region.
KW Human cytochrome P450; amplification; PCR; primer: expression vector;
KW yeast NADPH-P450 reductase; safety: fusion protein; metabolite:
KW carcinogen; mutagen; liver metabolism; ds.
OS Homo sapiens.
XX EP644267-A.
XX 22-MAR-1995.
XX 20-JUL-1994: 94EP-0111298.
XX 21-JUL-1993: 93JP-0180246.
XX 20-JUL-1993: 93JP-0201120.
XX 30-JUL-1993: 93JP-0208279.
XX (HATA/) HAYASHI K.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX WPI: 1995-115991/16.
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX P-PSDB: AAR72374.
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX Examples: Page 78-80; 124pp: English.
XX The nucleotide sequence of the cDNA coding region for the human
XX auxiliary cytochrome P450 species 2C19. The gene encodes a protein of
XX 490 amino acids. The cDNA was amplified by PCR using the primers
XX AA087769-89. The product was cloned into the yeast expression vectors
XX pAAH5N or pAHR5N to produce the vectors p2C19 for the expression of the
XX cytochrome P450 alone or p2C19R for co-expression with the yeast
XX NADPH-P450 reductase. The vectors are used in a method for evaluating the
XX safety of a chemical compound by reacting the chemical compound with
XX recombinantly produced human cytochrome P450 molecular species 1A2
XX (AA087714), 2C5 (AA087715), 2E1 (AA087716), or 3A4 (AA087717) or their


```

FH      Key:      Location/Qualifiers
FT      CD5      1..1443
PT      /*tag= a
PM      JP02072879-A.
PD      13-MAR-1990.
PE      09-SEP-1988: 88JP-0225955.
PF      09-SEP-1988: 88JP-0225955.
PG      (AGEN ) AGENCY OF IND SCI TECH.
PH      WPI: 1990-121045/16.
PI      P-PSDB; AAR04043.
PJ      Recombinant plasmid, for yeast for liver treatment -
PK      cytochrome P-450MP gene obtd. from yeast.
PL      etc.
PM      Disclosure: Fig 4: 9pp; Japanese.
PN      The P-450 produced from plasmid pHP6 in a yeast expression system
PO      preferably Saccharomyces AH22, can be used to treat human liver
PP      disorders, oxidising various chemical substances.
PQ      Sequence 1818 BP; 516 A; 412 C; 376 G; 514 T; 0 other:
PS
PT
PU
PV
PW
PX
PY
PZ
QA
QB
QC
QD
QE
QF
QG
QH
QI
QJ
QK
QL
QM
QN
QO
QP
QQ
QR
QS
QT
QU
QV
QW
QX
QY
QZ
RA
RB
RC
RD
RE
RF
RG
RH
RI
RJ
RK
RL
RM
RN
RO
RP
RQ
RR
RS
RT
RU
RV
RW
RX
RY
RZ
SA
SB
SC
SD
SE
SF
SG
SH
SI
SJ
SK
SL
SM
SN
SO

```

Query Match 83.3%; Score 1455; DB 11; Length 1818;
Best Local Similarity 92.3%; Pred. No. 0;
Matches 1574; Conservative 0; Mismatches 110; Indels 23; Gaps 3:

```

OY 57 tcaatctgagcagagctctggaggagaaactctctgccccactctctccca 116
DB 57 tcaatctgagcagagctctggaggagaaactctctgccccactctctccca 90
OY 117 gtgattgaaatctccatcagatcagatcagatcagatcagatcagatc 176
DB 117 gtgattgaaatctccatcagatcagatcagatcagatcagatcagatc 150
OY 177 tcaaaatctatgcccctgctctctctctctctctctctctctctctct 236
DB 177 tcaaaatctatgcccctgctctctctctctctctctctctctctctct 210
OY 237 catgatgatgaagtgtgagagagccctgattgattgagagagagttct 296
DB 237 catgatgatgaagtgtgagagagccctgattgattgagagagagttct 270
OY 297 ggcattctccatctgctgctgctgctgctgctgctgctgctgctgct 356
DB 297 ggcattctccatctgctgctgctgctgctgctgctgctgctgctgct 330
OY 357 aagagatgagagagagatcgccgtctctctctctctctctctctctct 416
DB 357 aagagatgagagagatcgccgtctctctctctctctctctctctct 390
OY 417 aagagagatgagagagatcgctgctgctgctgctgctgctgctgct 476
DB 417 aagagagatgagagagatcgctgctgctgctgctgctgctgctgct 450
OY 477 aaaaacagagctccatctgctgctgctgctgctgctgctgctgctg 536
DB 477 aaaaacagagctccatctgctgctgctgctgctgctgctgctgctg 510
OY 537 atctctccatctatctccagagagctgctgctgctgctgctgctgct 596
DB 537 atctctccatctatctccagagagctgctgctgctgctgctgctgct 570
OY 597 atggaadaattgaatgaacatcagagagtgagagagagagagagagag 656
DB 597 atggaadaattgaatgaacatcagagagtgagagagagagagagagag 630

```

```

OY 657 aatttccactatcattatatttccggagaaaccataaaaaatttctaaaa 716
DB 657 aatttccactatcattatatttccggagaaaccataaaaaatttctaaaa 690
OY 717 gctttatgagaagtgatatttggagaagataaaagaaacacacacacac 776
DB 717 gctttatgagaagtgatatttggagaagataaaagaaacacacacacac 750
OY 777 acaacccctcgagacttatgatctctctgacataaaatgagagaagaa 836
DB 777 acaacccctcgagacttatgatctctctgacataaaatgagagaagaa 804
OY 837 caacagctgaattcactattgaaacattgtaacacagctgacgtactt 896
DB 837 caacagctgaattcactattgaaacattgtaacacagctgacgtactt 864
OY 897 ggaacagagaaacaaacacacacacacacacacacacacacacacac 956
DB 897 ggaacagagaaacaaacacacacacacacacacacacacacacacac 924
OY 957 gagctcacagctcaagctcagagagagatgaacgtatcatttggaga 1016
DB 957 gagctcacagctcaagctcagagagagatgaacgtatcatttggaga 984
OY 1017 tggatcagacagatcagacacacacacacacacacacacacacacac 1076
DB 1017 tggatcagacagatcagacacacacacacacacacacacacacacac 1044
OY 1077 taactcagactctccacacacacacacacacacacacacacacacac 1136
DB 1077 taactcagactctccacacacacacacacacacacacacacacacac 1104
OY 1137 aactacttaattccacagagacacacacacacacacacacacacacac 1196
DB 1137 aactacttaattccacagagacacacacacacacacacacacacacac 1164
OY 1197 aaaaagaatttccacacacacacacacacacacacacacacacacac 1256
DB 1197 aaaaagaatttccacacacacacacacacacacacacacacacacac 1224
OY 1257 aattttaagaagaatcactctcagtgcttctcagcagagaaacagga 1316
DB 1257 aattttaagaagaatcactctcagtgcttctcagcagagaaacagga 1284
OY 1317 gagggctgcccagatgagctgttttctcagcactctcatttccagaa 1376
DB 1317 gagggctgcccagatgagctgttttctcagcactctcatttccagaa 1344
OY 1377 ctgaatctctgattgacccaaagacactctcagcactctctgttcag 1436
DB 1377 ctgaatctctgattgacccaaagacactctcagcactctctgttcag 1404
OY 1437 tctatccagcctctcagatcagctctcagcactctcagcactctc 1496
DB 1437 tctatccagcctctcagatcagctctcagcactctcagcactctc 1464
OY 1497 gctgctctgctgctcagcactctcagcactctcagcactctcag 1549
DB 1497 gctgctctgctgctcagcactctcagcactctcagcactctcag 1524
OY 1550 tctgtaagcttctcagcagcagcagcagcagcagcagcagcagcag 1609
DB 1550 tctgtaagcttctcagcagcagcagcagcagcagcagcagcagcag 1584
OY 1610 cactcagcctccacta-----aaaaagtttccagtgcaaatatctct 1659
DB 1610 cactcagcctccacta-----aaaaagtttccagtgcaaatatctct 1644
OY 1660 cccatctctatagtttctcagcagcagcagcagcagcagcagcagc 1719
DB 1660 cccatctctatagtttctcagcagcagcagcagcagcagcagcagc 1704

```



```
Db 848 caacagcaacacccctgagatgctctctctctctctgctgagcagccagaggtccag 907
Oy 967 ctaagtcacaggaagagatgaagctgtctatttgcagaaacccgagcccttgcagcacy 1026
Db 908 ctaagtcacaggaagagatgaagctgtctatttgcagaaacccgagcccttgcagcag 957
Oy 1027 acagggccacatgcccctacacagatgctgtgtgagcagaggtccagagataacacgac 1086
Db 968 cagggccacatgcccctacacagatgctgtgtgagcagaggtccagagataacacgac 1027
Oy 1087 tcatcccacagcctgcccctacacagatgctgtgtgagcagaggtccagagataacacgac 1146
Db 1028 tcatcccacagcctgcccctacacagatgctgtgtgagcagaggtccagagataacacgac 1087
Oy 1147 tcccaagggcacaacacataaacttccctcacttctctgctgctacatgacacaaagaat 1206
Db 1088 tcccaagggcacaacacataaacttccctcacttctctgctgctacatgacacaaagaat 1147
Oy 1207 tcccaagggcacaacacataaacttccctcacttctctgctgctacatgacacaaagaat 1265
Db 1148 tcccaagggcacaacacataaacttccctcacttctctgctgctacatgacacaaagaat 1207
Oy 1267 aaagttaactactctgcttctctcctcagcagaaacagcagatttctgtgtgagagggcctgg 1326
Db 1208 aaagttaactactctgcttctctcctcagcagaaacagcagatttctgtgtgagagggcctgg 1267
Oy 1327 ccgcagtgagctgttttatttctcagcttctcatttctcagaaacttcaactgaacttc 1386
Db 1268 ccgcagtgagctgttttatttctcagcttctcatttctcagaaacttcaactgaacttc 1327
Oy 1387 tgaatgacccaaagacacttgacacacactctgtgtcaatgatttctgtctccgc 1446
Db 1328 tgaatgacccaaagacacttgacacacactctgtgtcaatgatttctgtctccgc 1387
Oy 1447 ccttctcagctgtcttctcattctctgctgagaaacagcagatgctgtgtgagcctcg 1506
Db 1388 ccttctcagctgtcttctcattctctgctgagaaacagcagatgctgtgtgagcctcg 1447
Oy 1507 tgct 1510
Db 1448 gtct 1451

RESULT 15
AAV12758
ID AAV12758 standard; cDNA; 1447 bp.
XX 08-MAY-1998
XX AAV12758;
XX 08-MAY-1998 (first entry)
XX Cytochrome P450 2C19 mouse gene.
XX Cytochrome P450 2C19; human; uninterrupted metabolism: omeprazole;
XX diazepam; imipramine; miltin; ss.
XX Homo sapiens.
XX 4.1419
XX Location/Qualifiers
XX CDS
XX /cag- a
XX /transl_except- (pos: 382..384, aa:Asn)
XX /transl_except- (pos: 496..498, aa:Asn)
XX /transl_except- (pos: 520..522, aa:Asn)
XX /transl_except- (pos: 523..525, aa:Asn)
XX /transl_except- (pos: 586..588, aa:Asn)
XX /transl_except- (pos: 700..702, aa:Asn)
XX /transl_except- (pos: 772..774, aa:Asn)
XX /transl_except- (pos: 778..780, aa:Asn)
XX /transl_except- (pos: 916..918, aa:Asn)
XX /transl_except- (pos: 964..966, aa:Asn)
XX /transl_except- (pos: 1012..1014, aa:Asn)
XX /transl_except- (pos: 1305..1308, aa:Asn)
XX /transl_except- (pos: 1396..1398, aa:Asn)
XX JPI0033165-A.
XX 10-FEB-1998.
XX 23-JUL-1996; 56JP-0193015.
XX 23-JUL-1996; 56JP-0193015.
XX (SHIO ) SHIONOGI & CO LTD.
XX WPI: 1958-172087/16.
XX P-PSDB; AAN41160.
XX Modified cytochrome P450 2C19 from human liver - useful for
XX uninterrupted metabolism of therapeutics, e.g. omeprazole
XX Claim 4: Page 9-10; 15pp; Japanese.
XX This sequence represents the modified version of the human liver derived
XX cytochrome P450 2C19 gene of the invention. The modifications comprise
XX a substitution of the second amino acid with Ala and a deletion
XX of amino acid residues 3-20 of the wild type. Modified cytochrome
XX P450 2C19 is useful for: uninterrupted metabolism of therapeutics,
XX e.g. omeprazole, diazepam and imipramine in humans.
XX Sequence 1447 BP: 410 A: 341 C: 322 G: 374 T: 0 other:
Query Match 81.0%; Score 1413.6; DB 19; Length 1447;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 142; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Oy 67 gacagctctggagggagaaactctctctgcccactctctccacagatggaa 126
Db 11 gcaactctctggagggagaaactctctctgcccactctctccacagatggaa 70
Oy 127 atactctacagatattgaagatgacgaacactcttaaccaactctcaaaatct 186
Db 71 atactctacagatattgaagatgacgaacactcttaaccaactctcaaaatct 130
Oy 187 atggcctctgtctcactctgtatttggcctggaacgcatggtgctgcatgatg 246
Db 131 atggcctctgtctcactctgtatttggcctggaacgcatggtgctgcatgatg 190
Oy 247 aagtgtgaggaagccctctgattgattcttgagagagagatttctgagagagccattcc 306
Db 191 aagtgtgaggaagccctctgattgattcttgagagagagatttctgagagagccattcc 250
Oy 307 cactgctgaaagagcctaacagagagatttgaatcgttttcagcaatgagaaagatgga 366
Db 251 cactgctgaaagagcctaacagagagatttgaatcgttttcagcaatgagaaagatgga 310
Oy 367 aggaagattcgcctctctcctcctcctgacgtgcggaatttgggagggagagga 426
Db 311 aggaagattcgcctctctcctcctcctgacgtgcggaatttgggagggagagga 370
Oy 427 ttgagagacgttctcagagaaacccctcctctgtgagagatttgaagaaacccaagg 486
Db 371 ttgagagacgttctcagagaaacccctcctctgtgagagatttgaagaaacccaagg 430
Oy 487 cttaacccctgaccccttctcctcctgagcgtgctcctcgaatgctgctcca 546
Db 431 cttaacccctgaccccttctcctcctgagcgtgctcctcgaatgctgctcca 490
Oy 547 ttatttccagaacgttctcattataaagatcacgaacttcttaactgagggaaaaat 606
Db 491 ttatttccagaacgttctcattataaagatcacgaacttcttaactgagggaaaaat 550
Oy 607 tgaatgaaacacagagattgaagcccccctgagcagatgacgaatatttccca 666
Db 607 tgaatgaaacacagagattgaagcccccctgagcagatgacgaatatttccca 666
```

Db 551 tgaatgaatacctcagattgtaagcaccctcctggtacacagatgcaataattttccca 610
Oy 557 ctatcattgattttcccggaacccatacaaatctacttaaaacccctgtcttttaagg 726
Db 511 ctatcattgattttcccggaacccatacaaatctacttaaaacccctgtcttttaagg 670
Oy 727 aaagtgaatttttggaanaagaaacacccaagaatcgatggacatcaacaacccctc 786
Db 571 aaagtgaatttttggaanaagaaacacccaagaatcgatggacatcaacaacccctc 730
Oy 787 ggaactttatgatttctctcgtacataaataagaaagaaagcaaaacacacactctg 846
Db 731 ggaactttatgatttctctcgtacataaataagaaagaaagcaaaacacacactctg 790
Oy 847 aattcactatgaaaacttgatgaatcactcagctgacttacttggagctggagacagaga 906
Db 791 aattcactatgaaaacttgatgaatcactcagctgacttacttggagctggagacagaga 850
Oy 907 caacaagaacacccctggagataagctctctctctctctctctctctctctctctctc 966
Db 851 caacaagaacacccctggagataagctctctctctctctctctctctctctctctctc 910
Oy 967 ctaaagtccagagagatgaacgtgctcattggcagaaacccggagccctgcatcag 1026
Db 911 ctaaagtccagagagatgaacgtgctcattggcagaaacccggagccctgcatcag 970
Oy 1027 acaggggccaatgcctacacagatgctgtgtgacagaggctccagagatacatcgacc 1084
Db 971 acaggggccaatgcctacacagatgctgtgtgacagaggctccagagatacatcgacc 1030
Oy 1087 tcattccaccacagctcccatgacgtgactgtgacgttaaatccagaactaccctca 1146
Db 1031 tcattccaccacagctcccatgacgtgactgtgacgttaaatccagaactaccctca 1090
Oy 1147 ttcccaagggaacacacataataacttccctcactctctgtgctacatgacacaagaagt 1206
Db 1091 ttcccaagggaacacacataataacttccctcactctctgtgctacatgacacaagaagt 1150
Oy 1207 ttcccaaccacagagatttaacactgtcactctctctgtgactgaagtgaaatttaaga 1266
Db 1151 ttcccaaccacagagatttgaacctgtcactctctgtgactgaagtgaaatttttaaga 1210
Oy 1267 aaagtcaactactcattgctttctcagcaggaagaacgagatttgtgtggagaggccctgg 1326
Db 1211 aaagtcaactactcattgctttctcagcaggaagaacgagatttgtgtggagaggccctgg 1270
Oy 1327 ccgcatgagctgtttttatctctgacttcaatttaccagaactttaactgaaatctc 1386
Db 1271 ccgcatgagctgtttttatctctgacttcaatttaccagaactttaactgaaatctc 1330
Oy 1387 tgattgaccacaaggaacttgacacaactcctgtgtcaatgagatttctgtctccgc 1446
Db 1331 tgattgaccacaaggaacttgacacaactcctgtgtcaatgagatttctgtctccgc 1390
Oy 1447 ccttctacagctgtcttcaattctctgtctgaaagacagagatggtctgagctgct 1502
Db 1391 ccttctacagctgtcttcaattctctgtctgaaagacagagatggtctgagagct 1445

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 07:21:40 ; Search time 2723.34 Seconds
4889.382 Million CPU updates/sec

Title: US-09-763-292-1
Perfect score: 1746
Sequence: 1 cttcaatgctctttgtg.....acattattattataaga 1746

Scoring table: IDENTITY_MUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537288281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 01
Listing first 45 summaries

Database :

EST+
1: em_estfun+
2: em_esthum+
3: em_estin+
4: em_estom+
5: em_estpl+
6: em_estba+
7: em_estro+
8: em_estov+
9: em_hic+
10: em_hic2+
11: gb_est1+
12: gb_hic+
13: gb_gss+
14: em_gss_fun+
15: em_gss_hum+
16: em_gss_inv+
17: em_gss_pin+
18: em_gss_pro+
19: em_gss_vit+
20: em_gss_vit+
21: em_gss_other+*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	984.8	56.4	1990	12 AK008580	Mus muscu
2	966.6	55.4	1825	12 AK005017	Mus muscu
3	965	55.3	1870	12 AK008688	Mus muscu
4	788	45.1	1775	12 AK008512	Mus muscu
5	699.2	40.0	909	11 BG198899	RS118150
6	672.6	38.5	802	11 BG217233	RS118150
7	664	37.5	779	11 BG215664	RS118150
8	654	37.5	779	11 BG215664	RS118150
9	626	35.9	893	11 BG195327	RS115135
10	619	35.5	791	11 BG198890	RS118151
11	606.4	34.7	1017	10 A1529366	u167e10.y
12	592.4	33.9	910	11 BG198348	RS117505

ALIGNMENTS

```

RESULT 1
AK008580 1990 bp mRNA HTc 05-JUL-2001
LOCUS Mus musculus adult male small intestine cDNA, RIKEN full-length
DEFINITION enriched library, clone:2010318C06, full insert sequence.
ACCESSION AK008580
VERSION AK008580.1 GI:12842852
KEYWORDS CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male small intestine cDNA to
EMBL, GenBank, DDBJ, RIKEN full-length enriched mouse cDNA library
Accession: 2010318C06.
ORGANISM Mus musculus
Eukaryote; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1990)
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Methods in enzymology. 303, 19-44 (1999)
PUBMED 10319636
REFERENCE 2 (bases 1 to 1990)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome research. 10 (10), 1617-1630 (2000)
PUBMED 11042159
REFERENCE 3 (bases 1 to 1990)
AUTHORS Shihno,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,J., Nishi,K., Kitzumai,T., Tashiro,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

```


Oy	963	acacataaagtccagagaagaattgacagtgatcttgcagagaaacagagccctccatg	1022
Db	964		
Db	975	acagctaaagctccaggagaaaatttgacgcatgtgtcggcagacgacgaagcccatgcgatg	1034
Oy	1023	cacacagagggccacatgcctccacagatgctggtgtcagaggtgcagagatcacctc	1082
Db	1024		
Db	1035	cagacagacagagccatgcgcctctactcatgcgtcgtatgattctatgaggtccagagattcatt	1094
Oy	1083	gaactatccccacacagcctctccatgctgagctctgagctgagctgaattcagaacacac	1142
Db	1084		
Oy	1095	gaactatccccacacagcctctcccatgaagtgcacagtgatgacattttagatcagagacat	1154
Oy	1143	ctctcccccaagggcacaacacatataactcttccctcaactctgtgtctacatgacacaaa	1202
Db	1144		
Db	1155	ttcatcccccaagggacaaactgattatgatactatctgtcatcagctgtcgtatcagacagac	1214
Oy	1203	gaattcccccaacagagatglttgacctctgctctctcttctgagtgaggtgggaatttt	1262
Db	1204		
Db	1215	gagttgtcccaacccagagaaaattttgacctctgctccacatttctagatgagatggaaatttt	1274
Oy	1263	aagaaagaaactactctactcgtcttctccagcagagaaacaggtgtgtggtgagagggc	1322
Db	1264		
Db	1275	aagaaagaaactactcgtcttctctcaacaggaaacaaatcagatgtgtcagagagggc	1334
Oy	1323	ctgcacccctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1382
Db	1324		
Db	1335	ctgcacccctcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	1394
Oy	1383	ttcttgattgacccaaggaactctgacacacgctgtgtgtgtcgaatgatttgcctctgc	1442
Db	1384		
Db	1395	ctctcgtgttcacccaaggaacatagatgtatcccccatactcatttgatgtgcccctcggt	1454
Oy	1443	cgcgcctctatcagctgtgcttctctctctgctcgaaga	1481
Db	1444		
Db	1455	ctctcgtgttcacccactctgcttctctctctctcgaaaa	1493
RESULT	3		
AK008688	1870 bp	MRNA	05-JUL-2001
LOCUS			
DEFINITION	Mus musculus adult male stomach cDNA, RIKEN full-length enriched		
ACCESSION	Library clone:2210009K14, full insert sequence.		
VERSION	AK008688.1		
KEYWORDS	GI:12843035		
SOURCE	CAP trapper.		
	Mus musculus (strain:C57BL/6J) adult male stomach cDNA to mRNA,		
	clone:11b:RIKEN full-length enriched mouse cDNA library		
	clone:2210009K14.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.		
REFERENCE	Mammalian Ethology, 1970.		
AUTHORS	Carninci,P. and Hayashizaki,Y.		
TITLE	High-efficiency full-length cDNA cloning		
JOURNAL	Methods in enzymology. 303, 19-44 (1999)		
MEDLINE	93279253		
PUBLISHED	10349636		
REFERENCE	2 (bases 1 to 1870)		
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,		
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to		
JOURNAL	prepare full-length cDNA libraries for rapid discovery of new genes		
MEDLINE	20499374		
PUBLISHED	11042159		
REFERENCE	3 (bases 1 to 1870)		
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagao,S., Sasaki,M., Carninci,P.,		
TITLE	Kono,H., Akiyama,J., Ishii,T., Tsutsumi,T., Nishino,H., Harada,A.,		
JOURNAL	Yamanoto,P., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,		
MEDLINE	Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,		
PUBLISHED	Yonekeda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawaji,M.,		

TITLE Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
JOURNAL RIKEN integrated sequence analysis (RISA) system-384 format
MEDLINE sequencing pipeline with 384 multicapillary sequencer
REFERENCE Genome research, 10 (11), 1757-1771 (2000)
PUBMED 20530913

AUTHORS 4. (bases 1 to 1870)
TITLE The RIKEN Genome Exploration Research Group Phase II Team and the
JOURNAL FANTOM Consortium: cloning of a full-length mouse cDNA collection
REFERENCE Nucleic acid research, 28 (1), 409-495 (2000)
PUBMED 11076861

AUTHORS 5. (bases 1 to 1870)
 Adachi, J., Arakawa, K., Akahira, S., Akimura, T., Aono, R., Arai, A.,
 Banagaki, T., Bara, A., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F.,
 Iwatsuki, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kawai, J.,
 Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T.,
 Miyazaki, A., Nishi, K., Nemura, K., Numazaki, R., Ohno, M., Okazaki, Y.,
 Okubo, D., Shimizu, S., Shiohara, A., Shiohara, A., Shiohara, T.,
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,
 Tanaka, T., Terajima, Y., Toya, T., Yamamura, T., Yasunishi, Y.,
 Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-1-1, Suhei-cho, Tsurumi-ku, Yokohama,
 URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
 further details.

FEATURES cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 the preparation of this cDNA library. The cDNA was inserted into
 pSPL3 vector and subsequently enriched for full-length by cap-trapper. cDNA went
 through one round of normalization to Rot - 5.0 and subtraction to
 Rot - 20.0. Second strand cDNA was prepared with the primer adapter
 of sequence [5'

GAGAGAGAGATTCGAGTAAATTAATCCGCCCCCCCC 3'], cDNA was cleaved
 with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLiR.

FEATURES Location/Qualifiers
Source 1. 1870
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:1905087"
 /db_xref="MGI:1919553"
 /clone="2210009K14"
 /sex="male"
 /tissue="brain"
 /dev_stage="adult"
 /age="1427"
 /note="putative"
 /codon_start=3
 /protein_id="BA525834.1"
 /db_xref="GI:12843036"
 /translation="PKLRRNSRRNUPRPTPLPIGHLQDLKDIKSLRNFSEV
 TGPVTLTGHPAYVLTGZAYKATDHEGAGRGVFFNDFKRCGYVSSCR
 VNSVSTQNPFLKIDQEFLEENWETLSSPFLQCNFPAYDLTPQBRPLH
 KRFANFELSKVQHSIDINNPFDLFLIMQEKHNPKEFTICMLVFTDA
 KFAAGTETTLTLRLSLILLKLPYVAKVOEIDCVIGHRSPCHDRHSMPTDA
 VLHEORYIDLLTSLPHAVTRDFVREYLIPRGITVIASLISVLYDKDFNPKEFD
 PSHFDERGFKKSDYFFSTGKICVGEGLRAELFLITILQNLKSPVLDLK
 LDTTPVANGFASVPKFOICTP1"

CDS

polyA_signal 1851..1856
 /note="putative"
 polyA_site 1870
 /note="putative"
 BASE COUNT 538 a 411 c 377 g 544 t
 ORIGIN
 Query Match 55.3% Score 965; DB 12; Length 1870;
 Best Local Similarity 79.9%; Pred. No. 1,2e-252;
 Matches 1136; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
 Oy 59 antctggagagagctctggagagagaaacccctctgcccctctctctccctc 118
 Db 1 ACTATGGAGGAGAAATCTCAGAGAGAAACCTCTCTCTGCGCCACCTCTCC 57
 Oy 119 gattggaatctctccagatagatatttagagatgcagcaatcttaaccaatctctc 178
 Db 6 CATTGGAAATTTCTTCAGATAGNCTTAGGACATCAGCAATCTCTGAGGAATTTTC 127
 Oy 179 aaaaatctatgctc 238
 Db 12 AAAAGTCTATGGCCCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 187
 Oy 239 tgatatgaagtggtggaagagccctctctctctctctctctctctctctctctctctc 298
 Db 188 TGAATATGAGCTGTGAAGAGCCCTTACTGTATCTGAGAGAGCTTGTGGAGAGG 247
 Oy 299 ccaattcccaactggtcgaagagctaaacagagagatttggaatgctgttttcagcaatgaaa 358
 Db 248 AGTTTTTCCAGTGTGTGATAAGTTAAGAAAACTGCGGGGTGTCTTTTCAGCAGTGGGAG 307
 Oy 359 gagatggaagagattcggcgtttctctctctctctctctctctctctctctctctctc 418
 Db 308 GACCTGGAAAGATGAGAGCTTCTCACTCATGACTTGGAAACTTTGGATGGGTAG 367
 Oy 419 gagaacattgagaacagctgtcagaagagagccctctctctctctctctctctctctc 478
 Db 368 GAGCAGCATTTAGGACACCAATTTCAAGAGAGGCTGCTGTCTGTGGATGAATTGAGGAA 427
 Oy 479 aaccaagcttccactctgagatccactctctctctctctctctctctctctctctctc 538
 Db 428 AACCAAGGTGAACCTGTGTATCCACCTCTCATCTCTGCTCTCTCTCTCTCTCTCTCT 487
 Oy 539 ctgctcattatttccagaacagcttctctctctctctctctctctctctctctctctc 598
 Db 488 CAGCTCATGTGTTTCCAGAACTGTTTGTATATTAAGATCAGAAATCTTACCTTCTT 547
 Oy 599 gaaataattgaaatgaaacatcagagatgtgaagacccctctctctctctctctctc 658
 Db 548 AGACATCTTAATGAAATGTGAGATCTTGAGTTCCCTTGGATTCAGATTGCAATAA 607
 Oy 659 tttctccactatctgattttctccggagacccctatacaaatctacttaaaaccttgc 718
 Db 608 TTTCTCTGCTGTCTGATTTATCTACGAGAGACACAAATTTACATTAATAATTTTTC 667
 Oy 719 ttattgagaagtgatatttggagaagtaaaagagacccacacagatctgagatcaca 778
 Db 668 CTTTGCAGACATTAATCTTTTGTCAAAAGTAAACAGCATCAAGCTCACTGGATTAA 727
 Oy 779 caacctctggagcttattgattctctctctctctctctctctctctctctctctc 838
 Db 728 CAATCTCGGGACTTTATTGATTGTGTTCTGATCAAAATGGACAGAGAAAGCAACCC 787
 Oy 839 acattctgattctatgaaacttgaacttgaacttgaacttgaacttgaacttgaacttgc 898
 Db 788 AAACAGGAGTTTACTGTGAGAACTTGGTCTTACAGCATCTGATCTGTTTGGTCTGG 847
 Oy 899 gacagagacaac 958
 Db 848 GACAGAGACAAC 907
 Oy 959 gttcacagctaaagtcaggagagattgaacgtatgattggcagaaacacacacacac 1018

124	TGCTGAAGCAACCCAGAGGTACAGCTMAAGTCCAGAGAGATTGAAGCTGTGTATGSCA	183
1003	gaaacaggagcccttgatcagcagcaggagggacatagccctacacagatgctgtgacg	1062
184	GAACCGAGCCCTTCGATCGAAGACAGAGCCACACATGCCCTACACAGTGTGTGTGGTCC	243
1063	ccagagctcagagatcatalcagacctcattcccacagcctgcgccatcagtgacctgtg	1122
244	CGAGAGCTCCAGAGATACATTGACCTTCTCCCCACAGCCTGGCCCATGACGTGACCTGTG	303
1123	agcttaaatctcgaataacttactctatcccaggggcacacatataactctcaactt	1182
304	ACATTAAATTCAGAAAGATCTCTCTTCGAGGGGCACCAATTAATTTCCTCATCT	363
1183	ctctcgtcagcacaacagaagatttcccaaccagaagatgttgacctctgaccttc	1242
364	CTGTGCTACATGACACAAAGAATTTCACACCGAGATGTGTGACCTCATCATCTTC	423
1243	tggatagaggtggaaatttttgaagaagtaacttactcttcagtccttctcagaggaac	1302
424	TGGATGAGGTGGCATTTTATGAGAAATTAATACCTCTGCTTCACGACGAAAC	483
1303	ggatttgtagagagagggctctgcgcagatggagctgttttatctcagaccttatt	1362
484	GGATTGTGTGGGAGAGCCCTGGCCGCAATGGAGCTGTTTATTCCTGACCTCCATT	543
1363	tccgaacttttaacctgaactctctgattgcccacaaagaccttgacacactctgtg	1422
544	TACAGAACTTAACTCTGAACTGTGTGTGTCACAGACATCTTGACACACTCCAGTG	603
1423	tcaatgattctgtctgtgccccctctctcagctcagctctcatctctctcagaag	1482
604	TCAATGGATTGCTCTGTGCGCCGCTCTACACAGCTGTGCTCATTCCTGTCTGAAGA	663
1483	gcacagatgctggtctctctgtctgcttcctcagcagctctctctctctctctg	1535
664	GAGCAATGGCTTGGCTGCTGTGTGAGCTCCTGCACTCTTTCCTCTGGGGGTTA	723
1536	tccaaattcactatctgtagtctctctcttcgcgcctcatctcaaatcttcccctccc	1595
724	TGCATCTTCATCTGCTGTATGATGCTTTCTCACCTGATCATTCATATTTCTTCCT	783
1596	caaatctctatct	1614
784	GAGCACTAGTGACATCT	802

RESULT	7
LOCUS	BC567504 810 bp RNA
DEFINITION	602586239F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715312 5' mRNA sequence.
ACCESSION	BC567504
VERSION	BC567504.1
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Hakariya, S. and Ozolsa: Chordata: Craniata: Vertebrata: Euteleostomomammalia: Primates: Hominidae: Catarrhini: Homidae: Homo.
AUTHORS	(bases 1 to 819)
TITLE	NIH-MGC http://mac.ncbi.nih.gov/ .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: rcstra@nihl.nih.gov Tissue Resources: http://www.ncbi.nlm.nih.gov/ DNA Library Preparation: CLONTECH Laboratories, Inc. Clone Distribution: The I.M.A.G.E. Consortium (LLNL) Cloning Distribution: InCyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov/

Plate: LCM1561 row: m column: 09
High quality sequence stop: 810.
Location/Qualifiers
source
1. 819
/organism="Homo sapiens"
/clone="IMAGE:4715312"
/db_xref="taxon:9606"
/clone_lib="H1080 MCC17"
/notes="Organ: liver; Vector: pDR-LIB (Clontech); Site: 3'
SfiI (ggccatattgccc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGGCCATATTGCGC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGGCACATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.95
kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NMCC Library.
BASE COUNT 218 a 216 c 172 g 213 t
ORIGIN

Query Match 38.2% Score 667.4; DB 11; Length 819;
Best Local Similarity 92.4%; Pred. No. 2.2e-171;
Matches 759; Conservative 0; Mismatches 51; Indels 11; Gaps 5;
Oy 771 gacatcaaaccttcggacttattatgcttcttctgatacaaatggagaagaag 830
Db 1 GACATGACACACCTCAGGACTTATTGATGCTTCTCTGTAAGAAATGGAGAGGAG 830
Oy 831 caaaccaaaccttcgaattcaactctgtaaaacttgtaactcactgcagctactt 890
Db 6 CACACCAACCTCAGGACTTATTGATGCTTCTCTGTAAGAAATGGAGAGGAG 890
Oy 891 gggctgggag 950
Db 121 GGGCTGGGAG 950
Oy 951 caccagaggttcacgttaagctcagagagagagagagagagagagagagagag 1010
Db 181 CACCCAGAGGTTCACAGCTTAAGTCCAGGAGAGATTGACGTGTGATTGGCAGAACCG 240
Oy 1011 agcccttcgacgacagaggggagagagagagagagagagagagagagagag 1070
Db 241 AGCCCTTCGATGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070
Oy 1071 cagagatcaactcactcaccacagagagagagagagagagagagagagagag 1130
Db 301 CAGAGATACATTCAGCTTCTCCGACAGGCTGCCCATGCGAGTGCAGCTGACATTAA 350
Oy 1131 ttcagaaactcactcattcccaagggagagagagagagagagagagagagagag 1190
Db 361 TTCAGAAATATCATTCCTCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1190
Oy 1191 catgacaaag 1250
Db 421 CATGACAAAG 1250
Oy 1251 ggtggaatttttaa-gaaagtaactcactcactcactcactcagagagagagag 1309
Db 481 GGTGGCAATTTAAGGAAAGTAATACCTTCATGCTTCTCAGCAGGAGAAAGGAT 540
Oy 1310 tctggagaggggctggccgacgagagagagagagagagagagagagagagag 1369
Db 541 TGTGGAG 600
Oy 1370 cttaacactgaactctgagtgagagagagagagagagagagagagagagag 1429
Db 601 CTTTACCTGAAATCTGCTGTTGACCAAGAGAGAGAGAGAGAGAGAGAGAG 660
Oy 1430 attgcttctgagcgcctctatcagctgagcttcttctctctctctctctctct 1489
Db 181 ACCGAG 240

Db 661 ATTGCTCTGTGCGG-CCTTGACAGCTGTGCTTCTTCTTCTGCTGTGAAGAGCAGA 719
Oy 1490 tggctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1542
Db 720 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 778
Oy 1543 ttcactatctgagctgctgctgctgctgctgctgctgctgctgctgctgct 1583
Db 779 TTGCTATCTGTGTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 818
RESULT 8
BG215664 779 bp mRNA EST 21-APR-2001
LOCUS RST3335 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
DEFINITION BG215664
ACCESSION BG215664.1 GI:13741685
VERSION BG215664.1
KEYWORDS EST.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 779)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Krashoc,D., McElligott,K., Clark,S., Mays,R., Smith,E.,
Veloso,L., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducard,J.
TITLE Genome-wide Protein Expression Libraries using Random
Activation of Gene Expression
JOURNAL Nat. Biotechnol. 19 (5), 440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 351 9596
Email: scott@atersys.com
High quality sequence stop: 511.
FEATURES
Location 1-779
Source /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression', the
Nature Biotechnology, In Press. Note that even though the
method was used these sequence tags are not necessarily
expressed in H1080 under normal circumstances."
BASE COUNT 205 a 210 c 158 g 206 t
ORIGIN
Query Match 37.5% Score 654; DB 11; Length 779;
Best Local Similarity 91.3%; Pred. No. 1e-165;
Matches 707; Conservative 0; Mismatches 60; Indels 7; Gaps 1;
Oy 826 aaagcgaac 885
Db 1 AAAGCACAACCAACCACTTGAATTTACTATTGAAAGCTTGGAAACACCTGAGTTG 60
Oy 886 tacttgagctggag 945
Db 61 TGTGTGAGCTGGGAG 120
Oy 946 tgaacacacagagctcagcctcaaatccagagagagagagagagagagagagagag 1005
Db 121 TGAAGCAGCAGAGATCAGCTAAGTCCAGGAGAGATTGACGTGTGATTGGCAGAA 180
Oy 1006 accgagagccttcgacgacgagagagagagagagagagagagagagagagagag 1065
Db 181 ACCGAG 240

Df	421	TGTTTGACCCAAAGAACCTTGAACCACTTCAGTTGTGATCGATTGGATTGCTCTGTGCCTC 480
Oy	1447	ccttcatacagctggtgcattcctcgtctgaagaacacagatggcttggtcgtccctg 1506
Df	448	tttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1512
Df	481	CCTCTACCACCTGCTGCTACTTCTGCTGTGAAGAGACAGATGGCTGCTGCTGCTG 540
Oy	1507	tactcttcgcagctcctctctctctctcg-----tccaattctcacctctctgaagc 1559
Df	485	tttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 1565
Oy	541	TGAGTCCCTGCAGCTGTCTTCTCTGCGGCATATCCCATCTTTCTACTATCTGTAATGC 600
Oy	1560	ttctctgaccctgcattctcaattcttctctcccacaagatctcagtaaacatcaggct 1619
Df	601	TTTTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1625
Oy	1620	gccataaaagtctcagct-----ggagatactctgatctcccccatct 1667
Df	661	CCATTACGGAGGTTTCTATTGTTCAGTCGGACAATAATATCTGCTATTCTGCATCT 720
Oy	1668	ctataaagtacatttagtgccacataagtctgcatactctcatttgttgagttattaa 1727
Df	721	TGTGAACCACTGATTTGACTGGGCACATATGCTCTACTTATCTATATGTGAGTATTAA 780
Oy	1728	catatttat 1738
Df	781	TATGNTATAT 791

RESULT 11

A1529366
LOCUS
DEFINITION
P-450, complete cds (MOUSE); mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

FEATURES
source

EST 19-MAR-1999

Mus musculus
Mus musculus
Embryos; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

1 (bases 1 to 1017)

Maria.M., Hillier.L., Kucaba.T., Martin.J., Beck.C., Wyllie.T.,
Underwood.K., Steptoe.K., Theising.B., Allen.M., Bowers.Y., Person
.B., Swaller.T., Gibbons.M., Pepe.D., Harveys.N., Schurk.R., Ritter
.E., Kohm.S., Shih.I., Jackson.F., Cardenas.H., McCann.R.,
The Washington University Genome Center
The WashU-NCI Mouse EST Project 1999

Unpublished (1999)
Contact: Marra M./Aashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 286 1800
Fax: 314 286 1800
Email: mouse@wustl.wustl.edu
This clone is available royalty-free through LMLN : contact the
IMAGE Consortium (info@image.lmln.gov) for further information.
MG1:971822
Seq primer: custom primer used
High quality: yes
Location/Qualifiers
organism="Mus musculus"
strain="C57BL/
db_xref="taxon:10090"
clone="IMAGE:1887498"
clone_lib="Sugano mouse liver mlia"
sex="female"
dev_stage="adult"
tag="phage10"

(note)-Organ: liver; Vector: pME18S-FU3; Site_1: Drail1
(CAGCTGTGTC); Site_2: Drail1 (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
IATGTCGCCCTTTTITTTTTTTTTTII double-stranded cDNA was


```

|||||
185 AGAAGCGAGGCGCCCTGCATGCGAGAGAGGCGACATGCGCTACACAGATGCTGGTG 244
Oy 1062 cagcaggtccagagatacatcctcctcccccagcctgcccctcagtcagtcgt 1121
Db 245 CAGCAGGTCCAGAGATACATGACCTCTCTCCCAACAGCCTGCCCATGAGCTGT 304
Oy 1122 gaagtttaattcagaactcctctctcccaaggagcagacacattatcctcact 1181
Db 305 GACATTAATTCAGAACTATCTCATTCGACAGGACACATTAATTTCCCTGACT 364
Oy 1182 tctgtctacatcagacagaatttcccaaccagagatgtttgacctcgtcacttt 1241
Db 365 TCTGTGCTACATGACACAAAGAAATTTCCCAACCCAGAGATGTTTGACCTCATCATTT 424
Oy 1242 ctgagtagagtggaatttgaagaagaaactcctcctctctcagcagaaaa 1301
Db 425 CTGGATGAGGTGGCAATTTTGAAGAAGTAATTAATTCATGCTTTCTCAGCAGAAA 484
Oy 1302 cgaatttctcagagagagcctgcccagagagagcttttttttttttttttttt 1361
Db 485 CGAATTTCTGTGGAGAGCCCTGGCGGAGTGGAGCTGTTTTATTCCTGACCTCAT 544
Oy 1362 ttacagaactttaaactgaattctctgattgaccacaaagaccttgacacactcgtt 1421
Db 545 TTACAGAACTTAACCTGAAATCTCTGTTGACCCAAAGAACCTTGACACCACTCAGNT 604
Oy 1422 gtaaatgattgttctctccgcctctctatcagctgtgtctcctctctctgaaga 1481
Db 605 GTCAATGATTTGCTCTGCTGCGCCTTCTACACGCTGCTTCATTCCTGCTGAGG- 663
Oy 1482 agcagatggtctgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 1541
Db 664 AGAACACATAGCCTGCTGCTGTTGTGAGTCCCTGAAATCTCTTTTGGGGCAITATC 723
Oy 1542 tttaactatgtagctctctctgacctcagcgcctcctcacttttccctc 1592
Db 724 ACTTGACTATCTGATGCTGCTTTTACCTGCTGCTTAAATTTTCCCTTC 774

RESULT 13
LOCUS BG205076 907 bp mRNA EST 21-APR-2001
DEFINITION R5724495 Athersys RAGE Library Homo sapiens cDNA. mRNA sequence.
ACCESSION BG205076
VERSION BG205076.1 GI:13726763
KEYWORDS EST.
SOURCE human sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 907)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Dahl,T., Thornton,M., Ramchandran,R., Whittington,J.,
Lerner,L., Kreshoc,D., McElligott,K., Clark,S., Mays,R., Saith,E.,
Velloso,N., Hess,J., Cothren,K., Lo,K., Offenbacher,J., Danzig,J.,
and Ducar,M.
TITLE Creation of Genome-wide Protein Expression Libraries using Random
JOURNAL Nucleic Acids Res 29(19):440 (2001) In press
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave. Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scai@atersys.com
High quality sequence stop: 354.
Loc: 907
Source:
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="Athersys RAGE Library"
/cell_line="HT1080"

```

```

/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 226 a 234 c 192 g 234 t 1 others
ORIGIN
Query Match 33.8% Score 589.8; DB 11; Length 907;
Best Local Similarity 86.9%; Pred. No. 3.8e-150;
Matches 720; Conservative 0; Mismatches 97; Indels 12; Gaps 6;
Oy 872 agggaaagcaaaacaa-cagctcgaattcactatgaaaacttggttaactcagcagc 880
Db 2 AGGAAAGCAACCAACCACTTGAATTTACTATTGAAGCTTGGAAACACTGCGT 61
Oy 881 tgaattacttgagctggagagagacacacacacacacacacacacacacacacac 940
Db 62 TCATCTGTTGGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
Oy 941 cctgtgaagcaccagaggtcacagctaaagtccagagagagatgaacgtgctggt 1000
Db 122 CTGCTGAAGCACCAGAGGTGACAGTAAAGTCCAGAGAGAGATTGAACGTGATGG 181
Oy 1001 cgaagcagagagcctcgtcagcagcagagggccacccctacacagagctgaggt 1060
Db 182 CAGAACCGAGACCCCTGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 241
Oy 1061 gcaagagtcagagatactcagctcctcctcagcagcctgcccacagcagctg 1120
Db 242 GCACAGGTCCAGAGATACATGACCTCTCTCCACAGCCTGCCCATGCGAGCTG 301
Oy 1121 tgacttaattcgaacttaactcctcctcagaggggcaacacacacacacacacac 1180
Db 302 TGACATTAATTTAGAACTATCTCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 361
Oy 1181 tctgtctacatgcacacagaatttccacccagagatgttgacctcgtcactt 1240
Db 362 TCTGTCTACATGACACAAAGAAATTTCCCAACCCAGAGATGTTGACCTCATCAT 421
Oy 1241 tctgagtagagtggaattttaaagaagtaactcctcctcctcagcagagaaa 1300
Db 422 TCTGATGAGGTGGCAATTTTAAAGAAAGTAATATCTTCCCTTCTCAGCAGAAA 481
Oy 1301 agagattgtatggagagagcctgcccagcagagatgttttttctcagcctcat 1360
Db 482 ACGGATTTGTGGGAGAGCCCTGGCGGATGAGAGCTGTTTTATTCCTGACCTCAT 541
Oy 1361 ttacagaactttaaactgaaatctctgtagcaccacaaagaccttgacacacactcg 1419
Db 542 TTTACAGAACTTTAACTGAAATCTCTGTTGACCCCAAGACCTTGACACCACTC 601
Oy 1420 ttgcaatgagattgtctgcccgcctctcag--ctgtgctcactcctgctg 1477
Db 602 TTGTATTTGGATTTGCTTTTCTCCGCCCTTCTACCGGCTGTGGCTTATTTCTCTG 661
Oy 1478 agagacagcagtggtctgctcctcctcgtcgtcctcctcagcctcctcctcgt 1535
Db 662 GAGAAGAACAAACCCGCTGGCTGTGGTGGCAATCCCTGACGCTCTCTTTCTTGG 721
Oy 1536 ----tccaaattcactatctgtagcttctctgacccgctcactcacttctcctt 1591
Db 722 CTATTCATCTTTACTATCTGGAAAGCCT-TTTCACCTGGAAATTCACAAATTTCT 780
Oy 1592 ccccaagatctaggaactcagcctcctcctcctcctcctcctcctcctcgt 1640
Db 781 -CCTTGAGATCTAAGGAGCTTCCACATCATTACCGGAGTGCCTATGT 828

RESULT 14
BG186520

```

QY	1281	atccctttctcagacagaaacagatctgctggagagagcctcgccccatcgagcgtg	1340
Db	427	ATGCGTTTCTCAGCAGAGAAACGGATTTGTGTGGGAGAGCGCTTGCCGCGCATGCGAGCTG	486
QY	1341	ttttttattccctgacctctttttacagaaacttacctgaaatctctgattgaccctaaag	1400
Db	480	TTTCTTATCTGAGCTCCATTTTACAGAACTTTAACTGAAATCTTGTAACCCAAAG	546
QY	1401	gaactctcctcagtg	1460
Db	547	AACTCTTGACACCACTTCAGGTGTCTCAATGGGTTCGCTCTTGCCGCCCTCTACCAAGCTG	606
QY	1461	tgctctctctctctctctcagaagacagaaatgcttggtcgtctctctgctctctctcagag	1520
Db	607	TGCTTCATCTGCTCTTGAGAA-AACAGATGGCTCTGCTGTCTGTGCATTCGCAATCTGCA	665
QY	1521	ctctctctctctctctg-----tcccaattctcactctgactgactctctctctctctctg	1573
Db	666	CTCTCTTTGCTGTGGGGCTATCCATCTTTAACTACTTGTAATGC-CTTGCTCAAGCTG	723
QY	1574	cattccacatttttcccttcccccaaatcattgaaactcagctccattcattcaaaaagtt	1633
Db	724	CATCTCACATTCCTCTCTTGAGAACTAGCAATTCACCCATTCAGGGAGTTTCTATG	783
QY	1634	tcactctcaaatatatactctctctctccactct	1669
Db	784	TTACTGTGCATTTCTGATCTACTGACTTTTACACT	819

RESULT 15	AV53206	567 bp	mRNA	EST	07-SEP-2000
LOCUS	AV53206	GLC Homo sapiens	CDNA clone	GLCIDF03.3	mRNA sequence.
DEFINITION	GLC				
ACCESSION	U00001				
VERSION	AV53206.1	GI:9274220			
KEYWORDS	EST				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Cnidaria; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 567) Glan, B., Li, N., Peng, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H., Zeng, L., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Ou, J., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Lu, F., Ren, S., Zhong, M., Lu, C., Yang, Y., Gu, Y., Chen, Z., and Han, Z.				
TITLE	Homo sapiens CDNA clone				
COMMENT	Unpublished (2000) Contact: Zeguan Han Chinese National Human Genome Center at Shanghai 201200 Shanghai, China Tel: 86-21-50801519(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.				
FEATURES	Location/Qualifiers				
source	1..667 /organism="Homo sapiens" /clone="GLC" /clone="GLCIDF03" /clone="lib=GLC" /tissue_type="corresponding non cancerous liver tissue" /dev_stage="Adult" /lab_host="SOLR" /note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"				
BASE COUNT	193 a	126 c	151 g	192 t	5 others
ORIGIN					

Query Match	33.5%	Score 585.2	DB 10	Length 667:
Best Local Similarity	92.1%	Pred. No. 6.3e-149;		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 11:06:25 ; Search time 5830.82 Seconds
(without alignments)
1200.021 Million cell updates/sec

Title: US-09-763-292-3

Perfect score: 323
Sequence: 1 cccctgaattctacacaa.....aaattcatatcatctttt 323

Scoring table: IDENTITY NYC

Gapop 10.0, Gapext 1.0

Searched: 22023303 seqs, 10831430700 residues 44046506

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 47 summaries

Database : Pending Patents, NA, Main.*

- 1: /cgn2_6/ptodata/2/pna/US05_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US05_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US080_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US095E_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US095F_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US095G_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US095H_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US095I_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US095J_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US095K_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US095L_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US095M_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US095N_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US095O_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US095P_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US095Q_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US095R_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US095S_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US095T_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US095U_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US6023_COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Length	ID	Description
1	323	100.0	323	30	US-09-763-292-3
2	323	100.0	323	55	US-60-226-176-759
3	323	100.0	323	56	US-60-233-468-759
4	323	100.0	323	64	US-60-313-371-759
5	310.4	96.1	572	50	US-60-170-373-1355
6	310.4	96.1	572	51	US-60-181-428-187
7	310.4	96.1	572	51	US-60-181-428-187
8	310.4	96.1	572	51	US-60-181-428-187
9	310.4	96.1	572	51	US-60-181-428-187
10	299.4	92.7	10097	54	US-60-213-795-55
11	299.4	92.7	10097	54	US-60-213-795-55
12	299.4	92.7	10097	54	US-60-213-795-55
13	284.2	88.0	3557	53	US-60-207-211-20
14	284.2	88.0	3557	53	US-60-207-211-20
15	284.2	88.0	3557	53	US-60-207-211-20
16	284.2	88.0	3557	53	US-60-207-211-20
17	276.8	85.7	8437	18	US-09-488-127-59
18	276.8	85.7	8437	18	US-09-488-127-59
19	276.8	85.7	8437	18	US-09-488-127-59
20	276.8	85.7	8437	18	US-09-488-127-59
21	275.2	85.2	685	52	US-60-194-110-15
22	275.2	85.2	685	52	US-60-194-110-15
23	275.2	85.2	685	52	US-60-194-110-15
24	275.2	85.2	685	52	US-60-194-110-15
25	275.2	85.2	685	52	US-60-194-110-15
26	275.2	85.2	685	52	US-60-194-110-15
27	275.2	85.2	685	52	US-60-194-110-15
28	275.2	85.2	685	52	US-60-194-110-15
29	275.2	85.2	685	52	US-60-194-110-15
30	275.2	85.2	685	52	US-60-194-110-15
31	275.2	85.2	685	52	US-60-194-110-15
32	275.2	85.2	685	52	US-60-194-110-15
33	275.2	85.2	685	52	US-60-194-110-15
34	275.2	85.2	685	52	US-60-194-110-15
35	275.2	85.2	685	52	US-60-194-110-15
36	275.2	85.2	685	52	US-60-194-110-15
37	275.2	85.2	685	52	US-60-194-110-15
38	275.2	85.2	685	52	US-60-194-110-15
39	275.2	85.2	685	52	US-60-194-110-15
40	275.2	85.2	685	52	US-60-194-110-15
41	275.2	85.2	685	52	US-60-194-110-15
42	275.2	85.2	685	52	US-60-194-110-15
43	275.2	85.2	685	52	US-60-194-110-15

c 32 263 81.4 56476 57 US-60-245-227-19
c 33 263 81.4 56476 58 US-60-258-278-70
c 34 263 81.4 56476 59 US-60-248-572-73
c 35 254 78.8 38319 57 US-60-182-895-140
c 36 254 78.8 38319 57 US-60-182-895-140
c 37 250.4 77.5 495 49 US-60-162-247-1064
c 38 223.4 69.2 13925 57 US-60-248-498-60
c 39 223.4 69.2 13925 57 US-60-248-542-61
c 40 223.4 69.2 15443 54 US-60-213-795-28
c 41 223.4 69.2 36191 57 US-60-248-498-23
c 42 223.4 69.2 36584 57 US-60-245-227-10
c 43 223 69.0 1001 26 US-09-671-317-58
c 44 208.8 64.6 567 22 US-60-198-818-180
c 45 208.8 64.6 567 52 US-60-198-818-181

ALIGNMENTS

RESULT 1
US-09-763-292-3
: Sequence 3, Application US/09763292
: BEST LOCAL SIMILARITY 100.0%; Pred. No. 7.4e-90;
: Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
: GENERAL INFORMATION:
: APPLICANT: SmithKline Medical
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: Primers
: CURRENT APPLICATION NUMBER: US/09/763,292
: CURRENT FILING DATE: 2000-02-21
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 3
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: OTHER INFORMATION: GB:HUM2C9X05
US-09-763-292-3

Query Match 100.0%; Score 323; DB 30; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttaatttg 60
Db 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttaatttg 60
Oy 61 tcttatacagctaaagtccaggaagattgaacgtgattggcagaaccggagccct 120
Db 61 tcttatacagctaaagtccaggaagattgaacgtgattggcagaaccggagccct 120
Oy 121 gcatgcaagcagggccacatgcccatacagatgctgtgacgaggtccagaagt 180
Db 121 gcatgcaagcagggccacatgcccatacagatgctgtgacgaggtccagaagt 180
Oy 181 acctgaaccttcccacacagctgcccatacagatgctgtgacgaggtccagaagt 240
Db 181 acctgaaccttcccacacagctgcccatacagatgctgtgacgaggtccagaagt 240
Oy 241 acctatcattcccaggaagttgtttcttctacacgtgacatcatttttcgaagt 300
Db 241 acctatcattcccaggaagttgtttcttctacacgtgacatcatttttcgaagt 300
Oy 301 cccaaattcatatcatctttt 323
Db 301 cccaaattcatatcatctttt 323

RESULT 2
US-60-226-176-759
: Sequence 759, Application US/60226176
: BEST LOCAL SIMILARITY 100.0%; Pred. No. 7.4e-90;
: Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
: GENERAL INFORMATION:
: APPLICANT: Malsen, Gareth
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: Primers
: CURRENT APPLICATION NUMBER: US/60/233,468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 248
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: Misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-233-468-759

APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GB-0013-1 P
: CURRENT APPLICATION NUMBER: US/60/226,176
: CURRENT FILING DATE: 2000-08-16
: NUMBER OF SEQ ID NOS: 2447
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: Misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-226-176-759

Query Match 100.0%; Score 323; DB 55; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Oy 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttaatttg 60
Db 1 cccctgaatgctacacaaatgcccattttctctttccatcagtttttaatttg 60
Oy 61 tcttatacagctaaagtccaggaagattgaacgtgattggcagaaccggagccct 120
Db 61 tcttatacagctaaagtccaggaagattgaacgtgattggcagaaccggagccct 120
Oy 121 gcatgcaagcagggccacatgcccatacagatgctgtgacgaggtccagaagt 180
Db 121 gcatgcaagcagggccacatgcccatacagatgctgtgacgaggtccagaagt 180
Oy 181 acctgaaccttcccacacagctgcccatacagatgctgtgacgaggtccagaagt 240
Db 181 acctgaaccttcccacacagctgcccatacagatgctgtgacgaggtccagaagt 240
Oy 241 acctatcattcccaggaagttgtttcttctacacgtgacatcatttttcgaagt 300
Db 241 acctatcattcccaggaagttgtttcttctacacgtgacatcatttttcgaagt 300
Oy 301 cccaaattcatatcatctttt 323
Db 301 cccaaattcatatcatctttt 323

RESULT 3
US-60-233-468-759
: Sequence 759, Application US/60233468
: BEST LOCAL SIMILARITY 100.0%; Pred. No. 7.4e-90;
: Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0
: GENERAL INFORMATION:
: APPLICANT: Malsen, Gareth
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: Primers
: CURRENT APPLICATION NUMBER: US/60/233,468
: CURRENT FILING DATE: 2000-09-18
: NUMBER OF SEQ ID NOS: 248
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE: Misc-feature
: NAME/KEY: misc-feature
: OTHER INFORMATION: GB:HUM2C9X05
US-60-233-468-759

Query Match 100.0%; Score 323; DB 56; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;

Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 2;

Oy 1 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 60
Db 1 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 60

Oy 61 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 120
Db 61 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 120

Oy 121 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 180
Db 121 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 180

Oy 181 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 240
Db 181 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 240

Oy 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300
Db 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300

Oy 301 cccaaattcatagatcattttt 323
Db 301 cccaaattcatagatcattttt 323

RESULT 4
US-60-313-371-759
: Sequence 187, Application US/6031371
: GENERAL INFORMATION:
: APPLICANT: Ring, Ruijun Z.
: APPLICANT: Malsen, Gareth
: APPLICANT: Tomley, David
: APPLICANT: Morris, MacDonald
: TITLE OF INVENTION: Single Nucleotide Polymorphisms Associated With ADME Genes
: FILE REFERENCE: GR-0013-5 P
: CURRENT APPLICATION NUMBER: US/60313-371
: NUMBER OF SEQ ID NOS: 08-16
: SOFTWARE: PERL Program
: SEQ ID NO 759
: LENGTH: 323
: TYPE: DNA
: ORGANISM: Homo sapiens
: NAME/REF: misc.feature
: OTHER INFORMATION: GR:R02C9X05
US-60-313-371-759

Query Match 100.0%; Score 323; DB 64; Length 323;
Best Local Similarity 100.0%; Pred. No. 7.4e-90;
Matches 323; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 60
Db 1 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 60

Oy 61 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 120
Db 61 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 120

Oy 121 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 180
Db 121 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 180

Oy 181 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 240
Db 181 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 240

Oy 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300
Db 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300

Db 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300

Oy 301 cccaaattcatagatcattttt 323
Db 301 cccaaattcatagatcattttt 323

RESULT 5
US-60-170-373-1355/C
: Sequence 1355, Application US/60170373
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
: TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS AND USES
: FILE REFERENCE: THERDOP
: CURRENT APPLICATION NUMBER: US/60170-373
: NUMBER OF SEQ ID NOS: 198-12-13
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 1355
: LENGTH: 572
: TYPE: DNA
: ORGANISM: Human
US-60-170-373-1355

Query Match 96.1%; Score 310.4; DB 50; Length 572;
Best Local Similarity 96.4%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 60
Db 456 cccctgaattgctacaacaatgtgcattttctcttctccatcagttttacttg 397

Oy 61 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 120
Db 396 tcttatcagctaaagtccaggagagattgaactgtgtgattgacgaacccggagccct 337

Oy 121 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 180
Db 336 gcatgcaagacagagccacatgccctacacagatgctgtgtgacgaggtccagagat 277

Oy 181 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 240
Db 276 accttgactctcccccacagctgccctacagatgctgtgtgacatataattcagaa 217

Oy 241 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 300
Db 216 actatctctcccaagtaattttttctctacacagatgctgtgtgacatataattcagaa 157

Oy 301 cccaaattcatagatcattttt 323
Db 156 cccaaattcatagatcattttt 133

RESULT 6
US-60-181-428-187/C
: Sequence 187, Application US/60181428
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: FILE REFERENCE: CLO002258; US/60181-428
: CURRENT APPLICATION NUMBER: US/60181-428
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: FASTSEQ For Windows Version 4.0
: SEQ ID NO 187
: LENGTH: 572
: TYPE: DNA
: ORGANISM: HUMAN

US-60-181-428-187

Query Match 96.1% Score 310.4; DB 51; Length 572;
Best Local Similarity 99.4%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgctacacaaatggtccatttttctcttccatcagtttttactgtg 60
Db 456 CCCCTGAATTGCTACACAAATGGCCATTTTCTCTTTCATCATGTTTACTGTG 397
|||||
Qy 61 tcttatcagctaaagtcacgaagattgaacgtgtgattgagcaaacgggcccct 120
Db 396 TCTTATCAGCTAAAGTCCAGAGAGATTGACGTGTGATGGCAGAACCGAGCCCT 337
|||||
Qy 121 gcatgcaagcagagcagcagcagcagcagcagcagcagcagcagcagat 180
Db 326 GCATGCAGACAGAGCCACATGCCCTACACAGATGCTGTGTGTCACAGGTCCAGAT 277
|||||
Qy 181 acctgacctctcccacacagcctgccatgcagtgacctgtgacattaaattcagaa 240
Db 276 ACATGACCTTCTCCACACAGCCTGCCCATGCAGTGCATTAATTAATTCAGAA 217
|||||
Qy 241 acctatcctcccaagtaagtattttctctacacagcagcagcagcagcagat 300
Db 216 ACTATCTCATCCAGGTAAAGTTTGTCTCTACATGCACCTCACTGTTTTCGAAGT 157
|||||
Qy 301 -cccaaatccatagatcattttt 323
Db 156 CCCCAATTCATAGTATCATTTTT 133
|||||

RESULT 7
US-60-181-428-188/C
: Sequence 188, Application US/60181428
: GENERAL INFORMATION:
: APPLICANT: Bonazzi, Vivien
: TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL000227
: CURRENT APPLICATION NUMBER: US/60/181.428
: NUMBER OF SEQ ID NOS: 440
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 188
: LENGTH: 572
: TYPE: DNA
: ORGANISM: HUMAN
US-60-181-428-188

Query Match 96.1% Score 310.4; DB 51; Length 572;
Best Local Similarity 99.4%; Pred. No. 8.4e-86;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgctacacaaatggtccatttttctcttccatcagtttttactgtg 60
Db 456 CCCCTGAATTGCTACACAAATGGCCATTTTCTCTTTCATCATGTTTACTGTG 397
|||||
Qy 61 tcttatcagctaaagtcacgaagattgaacgtgtgattgagcaaacgggcccct 120
Db 396 TCTTATCAGCTAAAGTCCAGAGAGATTGACGTGTGATGGCAGAACCGAGCCCT 337
|||||
Qy 121 gcatgcaagcagagcagcagcagcagcagcagcagcagcagcagat 180
Db 326 GCATGCAGACAGAGCCACATGCCCTACACAGATGCTGTGTGTCACAGGTCCAGAT 277
|||||
Qy 181 acctgacctctcccacacagcctgccatgcagtgacctgtgacattaaattcagaa 240
Db 276 ACATGACCTTCTCCACACAGCCTGCCCATGCAGTGCATTAATTAATTCAGAA 217
|||||
Qy 241 acctatcctcccaagtaagtattttctctacacagcagcagcagcagcagat 300
Db 216 ACTATCTCATCCAGGTAAAGTTTGTCTCTACATGCACCTCACTGTTTTCGAAGT 157
|||||

US-60-248-498-247/C

Query Match 96.1% Score 310.4; DB 57; Length 51955;
Best Local Similarity 99.4%; Pred. No. 7.5e-85;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 cccctgaattgctacacaaatggtccatttttctcttccatcagtttttactgtg 60
Db 8041 CCCCTGAATTGCTACACAAATGGCCATTTTCTCTTTCATCATGTTTACTGTG 7982
|||||
Qy 61 tcttatcagctaaagtcacgaagattgaacgtgtgattgagcaaacgggcccct 120
Db 7981 TCTTATCAGCTAAAGTCCAGAGAGATTGACGTGTGATGGCAGAACCGAGCCCT 7922
|||||
Qy 121 gcatgcaagcagagcagcagcagcagcagcagcagcagcagcagat 180
Db 7921 GCATGCAGACAGAGCCACATGCCCTACACAGATGCTGTGTGTCACAGGTCCAGAT 7862
|||||
Qy 181 acctgacctctcccacacagcctgccatgcagtgacctgtgacattaaattcagaa 240
Db 7861 ACATGACCTTCTCCACACAGCCTGCCCATGCAGTGCATTAATTAATTCAGAA 7802
|||||
Qy 241 acctatcctcccaagtaagtattttctctacacagcagcagcagcagcagat 300
Db 7801 ACTATCTCATCCAGGTAAAGTTTGTCTCTACATGCACCTCACTGTTTTCGAAGT 7742
|||||
Qy 301 -cccaaatccatagatcattttt 323
Db 7741 CCCCAATTCATAGTATCATTTTT 7718
|||||

RESULT 9
US-60-245-227-9/C
: Sequence 9, Application US/60245227
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CL00087
: CURRENT APPLICATION NUMBER: US/60/245.227
: NUMBER OF SEQ ID NOS: 129

```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 9
: LENGTH: 225532
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)-(225532)
: OTHER INFORMATION: n = A,T,C or G
US-60-245-227-9

Query Match          96.1%  Score 310.4; DB 57; Length 225532;
Best Local Similarity 95.1%  Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

OY 1 cccctgaattgtctacacaaatgagccattttctcttccatcagttttacttg 60
DB 54498 CCCCTGAATTGCTACACAAATGAGCCATTTCCTCTTCCATCAGTTTACTTGTG 54439

OY 61 tcttatcagctaaagtcacgaagagattgaacgtgtgattgacgaacccggagccct 120
DB 54438 TCTTATCAGCTAAAGTCACGAAGAGATTGAACGTGTGATTGACGAACCCGGAGCCCT 54379

OY 121 gcatgcagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 179
DB 54378 GCATGCAGACAGACACACATG-CCTTACACAGATGCTGTGTGACGAGGTCCAGAGT 54319

OY 181 accttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 240
DB 54318 ACATTGACCTTCTCCACCAGCCTGCCATGCAGTGAACCTGTGACATTAATTCAGA 54259

OY 241 actatctatcccaaggaaagtgtttctcttccatcagctgaactcattttcagat 300
DB 54258 ACTATCTATCCCAAGGAAAGTGTTCCTCTTCCATCAGCTGAACCTCATTTCAGAT 54199

OY 301 -ccccaaattcatagatcatttt 323
DB 54198 CCCCAAAATTCATAGTATCATTTT 54175

RESULT 10
US-60-213-795-55
: Sequence 55; Application US/60213795
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: FILE REFERENCE: CL000705
: CURRENT FILING DATE: 2000-06-23
: NUMBER OF SEQ ID NOS: 267
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 55
: LENGTH: 10097
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)-(10097)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-795-56

Query Match          92.7%  Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%  Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;

OY 1 cccctgaattgtctacacaaatgagccattttctcttccatcagttttacttg 60
DB 5009 CCCCTGAATTGCTACACAAATGAGCCATTTCCTCTTCCATCAGTTTACTTGTG 5068

OY 61 tcttatcagctaaagtcacgaagagattgaacgtgtgattgacgaacccggagccct 120
DB 5059 TCTTATCAGCTAAAGTCACGAAGAGATTGAACGTGTGATTGACGAACCCGGAGCCCT 5128

OY 121 gcatgcagacagacacacatg-ccttacacagatgctgtgacgaggtccagaga 179
DB 5129 GCATGCAGACAGACACACATG-CCTTACACAGATGCTGTGTGACGAGGTCCAGAGA 5188

OY 180 tacttgaccttctccaccagcctgccatgcagtgacctgtgacattaaattcaga 239
DB 5189 TACTTGACCTTCTCCACCAGCCTGCCATGCAGTGACCTGTGACATTAAATTCAGA 5248

OY 240 actatctatcccaaggaaagtgtttctcttccatcagctgaactcattttcagag 299
DB 5249 ACTATCTATCCCAAGGAAAGTGTTCCTCTTCCATCAGCTGAACCTCATTTCAGAG 5308

OY 300 t-ccccaaattcatagatcatttt 323
DB 5309 TCCCCAAATTCATAGTATCATTTT 5333

US-60-213-795-56
: Sequence 56; Application US/60213795
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: FILE REFERENCE: CL000705
: CURRENT FILING DATE: 2000-06-23
: NUMBER OF SEQ ID NOS: 267
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 56
: LENGTH: 10097
: TYPE: DNA
: ORGANISM: HUMAN
: NAME/KEY: misc_feature
: LOCATION: (1)-(10097)
: OTHER INFORMATION: n = A,T,C or G
US-60-213-795-56
```

RESULT 12
US-60-213-795-57
Query Match 92.7%; Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000705
CURRENT APPLICATION NUMBER: US/60/213.795
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQUENCE ID NO 57
LENGTH: 10097
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: #isc_feature
LOCATION: 111 (10097)
ORIGIN INFORMATION: n = A, T, C or G
US-60-213-795-57

Query Match 92.7%; Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000705
CURRENT APPLICATION NUMBER: US/60/213.795
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQUENCE ID NO 57
LENGTH: 10097
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: #isc_feature
LOCATION: 111 (10097)
ORIGIN INFORMATION: n = A, T, C or G
US-60-213-795-57

Query Match 92.7%; Score 299.4; DB 54; Length 10097;
Best Local Similarity 99.1%; Pred. No. 9.3e-82;
Matches 322; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000705
CURRENT APPLICATION NUMBER: US/60/213.795
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQUENCE ID NO 57
LENGTH: 10097
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: #isc_feature
LOCATION: 111 (10097)
ORIGIN INFORMATION: n = A, T, C or G
US-60-213-795-57

Query Match 88.0%; Score 284.2; DB 53; Length 3557;
Best Local Similarity 97.9%; Pred. No. 3.1e-77;
Matches 320; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000583
CURRENT APPLICATION NUMBER: US/60/207.211
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQUENCE ID NO 21
LENGTH: 3557
TYPE: DNA
ORGANISM: HUMAN
US-60-207-211-21

Query Match 88.0%; Score 284.2; DB 53; Length 3557;
Best Local Similarity 97.9%; Pred. No. 3.1e-77;
Matches 320; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000583
CURRENT APPLICATION NUMBER: US/60/207.211
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQUENCE ID NO 21
LENGTH: 3557
TYPE: DNA
ORGANISM: HUMAN
US-60-207-211-21

QY 298 agt-cccccaattcagtagtattttt 323
||| ||||| ||||| ||||| ||||| |||||
Db 2326 agtcccccaattcagtagtattttt 2352

RESULT 15

US-60-207-211-22
: Sequence 22. Application US/60207211
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN PHASE I DRUG-METABOLIZING
: TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
: TITLE OF INVENTION: DRUG-METABOLIZING PHASE I PROTEINS, AND USES THEREOF
: FILE REFERENCE: CLO00583
: CURRENT FILING DATE: 2000-05-26
: NUMBER OF SEQ ID NOS: 74
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 22
: LENGTH: 3557
: TYPE: DNA
: ORGANISM: HUMAN
US-60-207-211-22

Query Match 88.04; Score 284.2; DB 53; Length 3557;
Basic Local Similarity 97.94; P-Value 3.1e-77;
Matches 320; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
QY 1 cccctgaatgctacacaaatgtccatttttctctctttccatcagtttttactatg 60
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2026 cccctgaatgctacacaaatgtccatttttctctctttccatcagtttttactatg 2085
QY 61 tcttatcagctaaagtcagagagatgaacgtgattggcagaaacc--ggagccc 118
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2086 tcttatcagctaaagtcagagagatgaacgtgattggcagaaaccggagccc 2145
QY 119 ctgcatgcaagcagagagcccatg-ccctacacagatgctggtgacagagctccaga 177
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2146 ctgcatgcaagcagagagcccatg-ccctacacagatgctggtgacagagctccaga 2205
QY 178 gataccttgacctctcccaaccagctgcccctgagtcgacctgagacattaaatca 237
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2206 gatacattgacctctcccaaccagctgcccctgagtcgacctgagacattaaatca 2265
QY 238 gaaactatctcattcccaagtgtaagttgttctctacactgcaactccatgttttcga 297
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2266 gaaactatctcattcccaagtgtaagttgttctctacactgcaactccatgttttcga 2325
QY 298 agt-cccccaattcagtagtattttt 323
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2326 agtcccccaattcagtagtattttt 2352

Search completed: April 19, 2002, 11:07:02
Job time: 13522 sec

Result No.	Score	Query Match	Length	DB	ID	Description
1	1739.6	99.6	1746	1	US-08-201-118-2	Sequence 2, Appli
2	1739.6	99.6	1746	2	US-08-238-8218-2	Sequence 2, Appli
3	1739.6	99.6	1746	5	PCT-US95-05744-2	Sequence 2, Appli
4	1534.6	87.9	1852	1	US-08-201-118-10	Sequence 10, Appl
5	1534.6	87.9	1852	2	US-08-238-8218-10	Sequence 10, Appl
6	1534.6	87.9	1852	5	PCT-US95-05744-10	Sequence 10, Appl
7	1529.8	87.6	1854	1	US-08-201-118-4	Sequence 4, Appli
8	1529.8	87.6	1854	2	US-08-238-8218-4	Sequence 4, Appli
9	1529.8	87.6	1854	5	PCT-US95-05744-4	Sequence 4, Appli
10	1436.4	82.3	1892	1	US-08-201-118-14	Sequence 14, Appl
11	1436.4	82.3	1892	2	US-08-238-8218-14	Sequence 14, Appl
12	1436.4	82.3	1892	5	PCT-US95-05744-14	Sequence 14, Appl
13	1373.6	78.7	1591	2	US-08-134-981E-3	Sequence 3, Appli
14	1276	73.1	1419	2	US-08-194-981E-4	Sequence 4, Appli
15	1364.6	72.4	2009	1	US-08-201-118-6	Sequence 6, Appli
16	1364.6	72.4	2009	2	US-08-238-8218-6	Sequence 6, Appli
17	1364.6	72.4	2009	5	PCT-US95-05744-6	Sequence 6, Appli
18	1364.6	72.4	2258	1	US-08-201-118-12	Sequence 12, Appl
19	1364.6	72.4	2258	2	US-08-238-8218-12	Sequence 12, Appl
20	1364.6	72.4	2258	5	PCT-US95-05744-12	Sequence 12, Appl
21	1170	67.0	1829	1	US-08-201-118-8	Sequence 8, Appli
22	1170	67.0	1829	2	US-08-238-8218-8	Sequence 8, Appli
23	1170	67.0	1829	5	PCT-US95-05744-8	Sequence 8, Appli
24	671.4	35.0	1829	2	US-08-194-981E-2	Sequence 2, Appli
25	671.4	35.0	1419	2	US-08-134-981E-2	Sequence 2, Appli
26	468	26.8	1737	2	US-08-750-703-2	Sequence 2, Appli
27	465.6	26.7	1740	2	US-08-750-703-1	Sequence 1, Appli

Query Match 99.6% Score 1739.6 DB 1: Length 1746;
Best Local Similarity 99.8% Pred. No. 0;
Matches 1742: Conservative 0; Mismatches 4; Indels 0; Gaps 0;

[illegible]

[illegible]

DB	1621	CATTAAAAAGTTTCACGTGCAAAATATATCTGCTATTCCCCATACCTATATATAGTTAC	1680
OY	1681	ATGAGTGCACAATATGATTAAGCTGCTATATGAGTATGATTAACCATATATATTA	1740
DB	1681	ATGAGTGCACAATATGATTAAGCTGCTATATGAGTATGATTAACCATATATATTA	1740
OY	1741	AATAGA	1746
DB	1741	ATNGA	1746

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

[illegible]

```

1141 ACCTCATCCCAAGGACACACATATATCTCCCTCCTCCTCTGCTGCTACATGACACA 1200
1201 agaattttcccaaccagagattgttgaccctcgtcacttctctgtagaggtgaaatt 1260
1201 AGAATTTCACACCCAGACATGTTTGACCTGCTCCTCTGCTGCTGCTGCTGCTGCTG 1260
1261 ttgagaagaatgacttctcagcttctcagcagagaacagagattgtgagagagag 1320
1261 TTAGAAGAATGACTACTCTGCTCTCTCAGAGAGAAACGATTTGCTGGAGAGG 1320
1321 gcttgagcagagagctgttttttctcagcctcttctacagaacttttaacctga 1380
1321 GCTTGAGCGGATGAGAGCTGTTTATCTCTGACCTCTATTTACAGAACTTTAACCTGA 1380
1381 aatctctgtagaccacaagacactctctgttcaatggaattgttctgt 1440
1381 AATCTCTGATGATGACCAAGACCTTGACACACTCTCTGTGTCATGATTTGCTCTG 1440
1441 tccgagccctctcagctctcctcctcctcagagaacagagagagagagagagag 1500
1441 TCCGAGCCCTCTATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
1501 ctctctgctgctcctcagcctctctctcctcctcctcctcctcctcctcctcctc 1560
1501 CTCTCTGCTGCTCCTGACGCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
1561 tctctgagcagctctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1620
1561 TCTCTGAGCAGCTCTCTCAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1620
1621 catcaaaagtctcctcagcagagagagagagagagagagagagagagagagagag 1680
1621 CATTAAGAGTTTCTGCTGCAATATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
1681 attgagtcacataatgctcagctctcctcctcctcctcctcctcctcctcctcctc 1740
1681 ATTAGAGTCACATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
1741 aataga 1746
1741 AATAGA 1746

RESULT 4
US-08-201-118-10
Patent No. 5786191
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
TITLE OF INVENTION: SUBFAMILY
NUMBER OF SEQUENCES: 44
CORRESPONDENCE ADDRESS:
RESIDENCE: Townsend and Townsend Kourile and Crew
STREET: 379 Linton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA: Release #1.0, Version #1.25
APPLICATION NUMBER: US/08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864, 962
```

```

FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: 'Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 3326-2400
FACSIMILE: (415) 3326-2400
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-201-118-10

Query Watch      87.9%; Score 1534.6; DB 1: Length 1852;
Best Local Similarity 92.9%; Pred. No. 0;
Matches 1637; Conservative 0; Mismatches 109; Indels 17; Gaps 2;

Cy 1 ctccaatgagcctcttggtgctgctgctgctgctgctgctgctgctgctgcttcca 60
Db 6 CTTCNAATGATATCTCTGTGGTCTGTGGTCTGTCTCATGTCTGTCTGTCTGTCTGT 65
Cy 51 tctgagcagagctctgagagagagagagagagagagagagagagagagagagagag 120
Db 56 TCTGAGAGAGAGCTCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 125
Cy 121 ttgagaatattctcacaatagatattgaagatgacagcaaatctcttaaccaatcttcaa 180
Db 126 TTGCAATATCTCTACAGATAGATTAAGGACATCAGCAAAATCTTAACCAATCTCTCAA 185
Cy 181 aaatctatgagcctctgttctcactctgtctgtctgtctgtctgtctgtctgtctgt 240
Db 186 AGGCTATGAGGCTGCTGTCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTG 245
Cy 241 aatatgaagtggtgagagagagagagagagagagagagagagagagagagagagag 300
Db 245 GATATGAGCAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 305
Cy 301 atttccactgagcagagagagagagagagagagagagagagagagagagagagagag 360
Db 306 TTTTCCCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 365
Cy 351 aatgagaagagatctgagcattctcctcctcctcctcctcctcctcctcctcctcct 420
Db 356 AATGGAAGAGAGATCTCCGCGCTTCTCTCTCATGACCTGCGGAAATTTTGGGATGGGA 425
Cy 421 ggaacatgagagagagagagagagagagagagagagagagagagagagagagagag 480
Db 426 GGAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 485
Cy 481 ccaagagctcagcctgagatccctctcctcctcctcctcctcctcctcctcctcctcct 540
Db 486 CCAGAGCTCACTCTGTGATCCCACTTTCATCTCTGGGCTGTCTCTCTGCTGCTGCTG 545
Cy 541 gctcctatttttccagaagagagagagagagagagagagagagagagagagagagag 600
Db 546 GCTCCATATTTTCCATAAAGCTTTTGATATTAAGATCAGCAATTTCTTAACCTTAATG 605
Cy 601 aaaaattgaagaacatcagagattgtaagcctcctcctcctcctcctcctcctcctcctcct 660
Db 606 AAAGTGTGATGAACATCAAGATTTTGACAGCCCTGGATCCAGATCTGCATATATT 665
Cy 661 ttccactcattgattatttcccgagagagagagagagagagagagagagagagagagag 720
Db 666 TTCTCTCTATCATGATGATTTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 725
Cy 721 ttatgagaagagatatttggagaagagagagagagagagagagagagagagagagagag 780
Db 726 TTATGAAAGTTATTTTGGAAAAAGTAAAGAACCAAGAGATCATATGAGCATGAGCA 785
```

QY	781	accctcggagcttattgatgtctctcgtacaaatgaggaagaaagcaaacacac	840
DB	786	ACCCTCAGGACTTATTGATGTCTCTGTATGAAATGAGGAAAGCAACACAC	845
QY	841	agtcgaattcaactatgaaaacttggtaactcactgcagctgacttcttgagctggaa	900
DB	846	CATCTGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	905
QY	901	cagagcagcaagcagcaacccctgaatactgctctctctctgctctaaaccccgag	960
DB	906	CAGAGACGACGACGACACACCTGAGATATGCTCTCTCTCTCTGCTGAACGACCGAGG	965
QY	961	tcaacgctaaagtccaggaagagattgaagctgtcatctggcgaacacggagcccttgc	1020
DB	966	TCACACGCTAAGGTCCAGGAGAGATTGAAGCTGTGATTGGCAGAACCGGAGCCCTCA	1025
QY	1021	tgccagcagggggccacaatgccctacacagatgctgttggtgcagaggtccagagata	1080
DB	1026	TCGAGGAGGACCCACATGCCCTCACAGATGCTGTGTGCAGAGGTGCAGATACG	1085
QY	1081	tcgaactcatcccaacagcctgcccaatgcagctgacccgttaaatctagaact	1140
DB	1086	TTGACCTCTCTCCCAACACAGCTGCCCATCGAGTGCAGCTGTGCATTTAAATTCAGAACT	1145
QY	1141	acctcatctccaggggcacaacacatataccttccctcactctgtgctacatgacaca	1200
DB	1146	ATCTCATATCCCAAGGGCACACCATATAATTTCCTGTACTCTGTGTACATGACACA	1205
QY	1201	agaatttcccaacccagagatgttgaccctcgtgcacttctggatgaagtggaatt	1260
DB	1206	AGAAATTTCCCAACCCAGAGTGTGTGACCTCATACTTCTCGATGAGGTGGCAAT	1265
QY	1261	ttgaagaagttaactactcatgctcttctcagcaggaagacgattgtgtggagag	1320
DB	1266	TTAGAAAAGTAAATACTTCATGCGCTTCTCAGCAGGAAGAACGGATTGTGTGGCAGAG	1325
QY	1321	gcctggcccgatggagctgttttattctctgacctcatlttacagaaacttbaactga	1380
DB	1326	CCCTGGCGCGCATGGAGCTGTGTTTATCTGTACACTGCATTTACAGAACTTAACTGA	1385
QY	1381	aattctgtatgcccaacagacactgcgaacactcgtgtgacacactgtgactgtgtc	1440
DB	1386	AATCTGTGTGTACCAAGACACCTTGACACACTCCAGTGTCAATGGATTGTGCTCTG	1445
QY	1441	tcacgctctctacagctgtgtctcatctctcgtgaagaagcacaatgctcgtgag	1500
DB	1446	TCCGCGCCCTCTACACGCTGCTCTCATCTCTGCTGAGAGAGACAGATGGCCTGGCTG	1505
QY	1501	ctcctgtgctgcctcgcagctctctctctctgg-----tccaaatttcaactatcg	1553
DB	1506	CTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1555
QY	1554	tgatgctctctcagcctcatctcacaatttctctcccccagaactcaagtgaact	1613
DB	1566	TAATGCTCTTCTCAGCTGTCACTCAGATTTTCCCTCTCCCTGAACATCTAGTGAACAT	1625
QY	1614	cagctccattla-----aaaaagtttccctgtgcaataatactcgtctattccca	1663
DB	1626	CGAAGCTCAATTAGGAGAGTTTCTCATGTTTCATGTGCAATAATATCTCTATTCTCA	1685
QY	1664	tactctataacgttacatgtaggcacataatgctgatactgtctgaattgtaata	1723
DB	1686	TACTCTGTACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1745
QY	1724	taacatattattataacag	1746
DB	1746	TTATATGTGTTTATTAAATAGA	1768

RESULT 5
US-08-238-821B-10
: Sequence 10, Application US/08238821B

Patent No. 5912120

GENERAL INFORMATION:

APPLICANT: GOLDSTEIN, Joyce A.

APPLICANT: ROMKES-SPARKS, Marjorie

APPLICANT: DE MORAIS, Sonia M.F.

TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN

TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-

TITLE OF INVENTION: MEPIHENTHOIN METABOLISM

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

COUNTRY: California

STATE: CA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/238,821B

FILING DATE: 06-MAY-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/201,118

FILING DATE: 22-FEB-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/864,962

FILING DATE: 09-APR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Liebeschuetz, Joe

REGISTRATION NUMBER: 37,505

REFERENCE/DOCKET NUMBER: 15280-1921100US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-326-2420

TELEFAX: 650-326-2423

INFORMATION FOR SEQ ID NO.: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1852 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: Region

LOCATION: 1..10

DESCRIPTION: (note: *corresponds to positions -10 to -1

OTHER INFORMATION: for 55 of Figure 2.

US-08-238-821B-10

Query Match	87.9%	Score 134.6	DB 2	Length 1852
Best Local Similarity	92.9%	Prod. No. 0		
Matches 1937	Conservative	Alimatches 109	Indels	Gaps
Qy	1	cttcaatgagatcttttggatcttgctgctgctctccatgcttggctctctcttccaa	60	
Db	6	cttcaatgagatcttcttggctgcttctgctctgctctcatgtttggcttgcctttccac	65	
Qy	61	cttggagagagagctctggagagagagaaactcctctctggcgcacactctctccacgaa	120	
Db	66	ctctggagacagagactctggagagagaaactcctctctggcgcacactctctccacgaa	125	
Qy	121	ttggaaatctctacagatagatattgaagatctgcagaaatccttaacaaatctctcaa	180	
Db	126	ttggaaatctctacagatagatattgaagatctgcagaaatccttaacaaatctctcaa	185	
Qy	181	aaatctatgacatctctgcctctctatatttggcctgaacacgaatgagtgctgcatg	240	
Db	186	aggtctatggccccttgcttacctctgattttggcctgaacacccatagtgctgctgcatg	245	

Oy 541 gctccattatttcgaagaaacgtttcagattataaagatcacgaattcttcaactatgatg 600
Db 548 gctccattatttccataaacggttttgattataaagatcagcaatttcttaactaatg 607
Oy 601 aaaaatgaagaaacatcagatgtatgaagccccccttgatccagatcatgaataatt 650
Db 608 aaaaatgaagaaacatcagatgtatgaagccccccttgatccagatcatgaataatt 657
Oy 661 tccctcatcatgatatttatttccgggaacccatacaaaattacttaaaacttgcgt 720
Db 668 ttcctcatcatgatatttatttccgggaacccatacaaaattacttaaaacttgcgt 727
Oy 721 tttggaagagatattttggagaagtaaaagacacccaagaatcgatgacatcaaca 780
Db 728 tttggaagagatattttttgggaagtaaaagacacccaagaatcgaatcgaatcaaca 787
Oy 781 accctcggaacttatttattgcttcctcatcaaaatggagaagaaagcaaaccaac 840
Db 788 accctcggaacttatttattgcttcctcatcaaaatggagaagaaagcaaaccaac 847
Oy 841 agctcatcttcacatcaaaacttgaatcctcagcagcagcagcagcagcagcagc 900
Db 848 catctgaatttacttattgaagccttggaaacactcagctgacttcttggagctgga 907
Oy 901 cagagacaacagcaacacccctgagatgcttctctctctctctctctctctctct 950
Db 908 cagagacaacagcaacacccctgagatgcttctctctctctctctctctctctct 957
Oy 961 tccagctaaagtcagagagatgagacgtgctatggcagaaacccgagccctctga 1020
Db 968 tccagctaaagtcagagagatgagacgtgctatggcagaaacccgagccctctga 1027
Oy 1021 tgcagcagagggccacatgcctcacaagatgctgtgacagcagcagcagcagcagc 1080
Db 1028 tgcagcagagggccacatgcctcacaagatgctgtgacagcagcagcagcagcagc 1087
Oy 1081 tgcagcagcagggccacatgcctcacaagatgctgtgacagcagcagcagcagcagc 1140
Db 1088 tgcagcagcagggccacatgcctcacaagatgctgtgacagcagcagcagcagcagc 1147
Oy 1141 acctcatctcccaagggccacacatcaacttccctcacttctgtctacatgacaca 1200
Db 1148 atctcatctcccaagggccacacatcaacttccctcacttctgtctacatgacaca 1207
Oy 1201 aagaatttcccaacccagagatgtttgacctctgctacatttcttgatgaaggtgaaat 1250
Db 1208 aagaatttcccaacccagagatgtttgacctctgctacatttcttgatgaaggtgaaat 1257
Oy 1261 ttaagaagaatcaactctcagccttctcagcagcaaaacagagattctgtgagagag 1320
Db 1268 ttaagaagaatcaactctcagccttctcagcagcaaaacagagattctgtgagagag 1327
Oy 1321 gctcggccgagatgagctgttttattctcgaacttcttattacagaactttaactga 1380
Db 1328 gctcggccgagatgagctgttttattctcgaacttcttattacagaactttaactga 1387
Oy 1381 aatctgtagtaccacaagagccttgacacacactctgtgtcaaggagatttgcctg 1440
Db 1388 aatctgtagtaccacaagagccttgacacacactctgtgtcaaggagatttgcctg 1447
Oy 1441 cccagcctctcagctgtctcagcagcagcagcagcagcagcagcagcagcagcagc 1500
Db 1448 cccagcctctcagctgtctcagcagcagcagcagcagcagcagcagcagcagcagc 1507
Oy 1501 ctcctgtgctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1553
Db 1508 ctcctgtgctcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1567
Oy 1554 tgaagcttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1613
Db 1568 tgaagcttctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1627

Oy 1614 caccctccatca-----aaaagttccactgctccaaatatactatctatccca 1663
Db 1628 caccctccatcaaacgaggttctctatgtttccactgtgcgaatatactctatctcca 1687
Oy 1664 tactcataatagttacatgagtcacataatgctgatactgtctaattgttgagtta 1723
Db 1688 tactctgttaacagttgcatgtgactgtgcataatgctctatacttacttaattgttgagtta 1747
Oy 1724 ttaacataattattataaalega 1746
Db 1748 ttaataattatttataaalega 1770

RESULT 10
US-08-201-118-14
: Sequence 14, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: INVENTOR: SPARKS, David L.
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/201.118
: FILING DATE: 09-APR-1992
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Liebeschuetz, Joe
: REGISTRATION NUMBER: 37,505
: REFERENCE/DOCKET NUMBER: 13280-192-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: RECOMBINATION: YES
US-08-201-118-14

Query Match 82.3% Score 1436.4 DB 1: Length 1892:
Best Local Similarity 86.5% Pred. No. 0:
Matches 1525: Conservative 0: Mismatches 221: Indels 17: Gaps 2:

Oy 1 cttcaatggatccttttgatgctcttgatctctctctctctctctctctctctctca 60
Db 38 cttcaatggatccttttgatgctcttgatctctctctctctctctctctctctctca 97
Oy 61 tctgagacagacactctgagagagaaactccctcctgagccacactctctcccaatga 120
Db 98 tctgagacagacactctgagagagaaactccctcctgagccacactctctcccaatga 157


```

APPLICATION NUMBER: US 07/864, 962
FILED: APR 1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
TELEPHONE/DOCKET NUMBER: 15280-19211005
TELECOMMUNICATION INFORMATION:
TELEFAX: (650) 326-2400
TELEPHONE: (650) 326-2422
INFORMATION FOR THE PUBLIC: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1892 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: YES
FEATURE:
FEATURE KEY: Region
LOCATION: 1..41
OTHER INFORMATION: /note= "Corresponds to positions -41 to-1
of Figure 2."
US-89-238-821B-14

Query Match      82.3%  Score 1436.4  DB 2:  Length 1892:
Best Local Similarity 86.5%:  Pred. No. 0:
Matches 1525:  Conservative 0:  Mismatches 221:  Indels 17:  Gaps

Oy      1  cttcaagaatcattcttgctctgctgctgctgctgctcctcctcattgctctctctca 60
Db      38  CTTCATCGATGCTCTNTGTGGCTCTGTCCTGTCCTGTCATGTGCTCTCTTCAC 97

Oy      61  ctggagacagagctctgggagagaaaactccctctggcccaactctctcccagtga 120
Db      96  TCTGGACAGAGCTCTGGGAGAGAAATCTCCCTCTGGCCACCTCTCTTCCCAATTA 157

Oy      121  tggaaatcctcagatagatataagagatcgacgaacatctcaacatctctca 180
Db      158  TCGAATAATCTCACATAGATNTTAGGACATCAGCAATCTCTTAACCATNTCTCA 217

Oy      181  aatctatggccttgctgaactctgatttggctggagacgatggctgcatg 240
Db      218  AAGCTATAGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277

Oy      241  gatataagatctgtagaagacgctgatgctgctgtagagagagatttctcgaa 300
Db      278  GATATGAACGAGTGAAGCAAGCCCTGATTGATGTTGAGAGAGAGATTCTTGAAG 337

Oy      301  attctccactggctgaaagcgtaacagagattgggaatcggtttcagcagtgaa 360
Db      338  HTTCTCCATCTGCTGTGAAGAGATACANAGGATNTGGATATCTTTCAGCATGGA 397

Oy      361  gatgaagatctgcgcgtttctcctcctcagtcgacgcgcggaatttggatgg 420
Db      398  GATGGAAGGAGATCGCGCCTTTCTCTCCCTCATGACGCTGGGGATTTTGGGATG 457

Oy      421  ggagcatggagacgctgttccaagagaaacgcgcctcttgaggagaatttgaaa 480
Db      458  GAGCATNTTGGACCTGTTCAGAGAGACCCCTGCTCTGTGGAGGATGTGAGAAA 517

Oy      481  cgaagcttcaccctggtaccctacattctatcgagcgcgtcctcctcagtgat 540
Db      518  CCAAGGCGCTACCCCTGTGATCCGCACTTTTCATCTCGGCGTGTGCTCGCTCA 577

Oy      541  gctccattatcttcagaaacgcttttcgattataaagatcagaatttcttaact 600
Db      578  CTCTCTTATTTTTCATTAACCTTGTATATTAAGATACAGATTTCTTACTTGT 637

Oy      601  aaaaatctgaagaaactcagatgtgaagcccccgcgcgcgcgcgcgcgcgcgc 660
Db      638  AAAAATTAATGAAGAAATCAGGATCTTGAGCCGCCGTCGATTCAGNTTCTCAAT 697

```

[illegible]

	841	GATTTACTATTGAAGCTTGGAAACCATGCACTGCATCTTGTTGGCTGTGGCAGAG	900
Db			
Oy	906	acacagcaaacacacacagatctctctctcctctgctgaagcacccagaggtacaa	965
Db			
Oy	901	ACGNACAGCACAAACCCTGAGATATGCTCTCTTCCTGCTGAAGCACCCAGAGGTACA	960
Db			
Oy	966	gctaagaicccaggaaagattgaagtcatcttgccaagaacaggacccttcgcagac	1025
Db			
Oy	961	CCTAAMGTCACAGAAGATATGAAGCTGTGATGTGCAGAACCGAGGCCCTGCTGCA	1020
Db			
Oy	1026	gacaagggcccacatgccctacacacagatgctgtgtccagagttccagagatacatgac	1085
Db			
Oy	1021	GACAGGACGCACATGCCCCCTACACAGATGCTGTGTGTCAGAGTCCAGAGATGCAT	1080
Db			
Oy	1086	ctcataccccaccagctgcgcccatgcagtagacctgtaogttaaactcagaacatacctc	1145
Db			
Oy	1081	CTCTGTCCCACACAGCTGCCCCCATGCTAGTGCATCTGTGACATAAATTCAGAACTATCTC	1140
Db			
Oy	1146	attcccaagggcaacacatatataactcccctcacttctgtgtcatcatgacacaaaga	1205
Db			
Oy	1141	ATTCACAGAGGNCANCCATATATTTCCCTGACTCTGTGCTACATGACACACAAAGA	1200
Db			
Oy	1206	tctcccaacacagagatttgagccctgcacttctctgagatgaagtgaaaatttaag	1265
Db			
Oy	1201	TTTCCCAACCCACAGATGTTTGCACCTCATCACTTCTGGATGAGGTGACAAATTTAAG	1260
Db			
Oy	1266	aaaaaglaactactctgcctcttctcaggagaaaacgagattgtgtgagagagggctg	1325
Db			
Oy	1261	AAMGTNAATACTTCGCTTTCTCAGCAGAAACGGATTTGTGTGGGAGAGCCCTG	1320
Db			
Oy	1326	gcccgatggagacttttttatcagatcctcttttagagactttaaacctgaataatc	1385
Db			
Oy	1386	ctgatgaccaccaagaacctkgacacacactctgtgtcaaagatttgtctctcccg	1445
Db			
Oy	1381	CTGTGTGACCAAMAAGACTTGCACACCACTCCAGTGTGTAATGGATTTGGCTCTGTGGCG	1440
Db			
Oy	1446	cctctctatcagctgtgcttcaattcctctctgagaagacagatgctgtgctgtcct	1505
Db			
Oy	1441	CCCTCTACACAGCTGTGCTTCATTTCTCTGTGAAGAAACAGCATGTGCTTGCT	1500
Db			
Oy	1506	gtctgtccctccagact	1563
Db			
Oy	1501	---CAGTCCCTTCGACGCTCTCTTCCCTCTGGGGCGATTATTCATCTTG	1560
Db			

RESULT 14

US-08-194-981E-4
: sequence 4, Application US/08194981E

Sequence #, Application 00/00134301
; Patent No. 5886157

; GENERAL INFORMATION

APPLICANT: GUENGEF

APPLICANT: GUO, Zuyu

APPLICANT: SANDHU, PU

; APPLICANT: GILLAM, Elizabeth

TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF

; TITLE OF INVENTION: HUMAN

TITLE OF INVENTION: CYTOCHROME P450

: NUMBER OF SEQUENCES: 68

; CORRESPONDENCE ADDRESS:

ADDRESSER: NEEDLE & ROSENBERG, P.C.

STREET: Sulte 1200, 127 Peachtree Street, NE

CITY: Atlanta

STATE: Georgia

; COUNTRY: USA

222: 30303-1811

; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; COMPILING: Detect In Release #1 0 Version #1 25

```

CURRENT APPLICATION DATA:									
APPLICATION NUMBER: US/08/194,981E									
FILING DATE: February 10, 1994									
CLASSIFICATION: 435									
ATTORNEY/AGENT INFORMATION:									
NAME: Elizabeth Selby									
REGISTRATION NUMBER: 38,298									
REFERENCE/DOCKET NUMBER: 2000,0022									
TELEPHONE: (404) 688-0770									
TELEFAX: (404) 688-9880									
INFORMATION FOR SEQ ID NO: 4:									
SEQUENCE CHARACTERISTICS:									
LENGTH: 1419 base pairs									
TYPE: nucleic acid									
TOPOLG: Single									
RANDOMNESS: High									
MOLECULE TYPE: DNA (genomic)									
HYPOTHETICAL: NO									
ANTI-SENSE: NO									
US-08-194-981E-4									
Query Match	73.1%	Score 1276:	DB 2:	Length 1419:					
Res. Locs. Similarity	92.0%	Pr. No. 0:							
Matches 1327:	Conservative	0:	Mismatches	85:	Indels	0:	Gaps	0:	
QY	67	gacagcgcctggagaggaacatccctctggcccccacatccctccacgtagatggaa	126						
DB	8	GACAAATCTCTGGACGAGGAAATCCCTCTGCGCCCACTCTCTCCAGTAGTATGGAA	67						
QY	127	atattctacgatgatataaagatgctgacgaataatccctaaacatctctcaaaagtct	186						
DB	58	ATATCTACATATGATTTTAGGACATACGAAATCTTTACCAATCTCTCAAAAGTCT	127						
QY	187	atgacgcctgactctgctgatttggcctggaacgcagtcagtcgtctcgtcgtgaaatg	246						
DB	128	ATGGCCCTCTGTCCTACCTCTGATTTTGGCTTGAACCCATAGTGTGTCATCGATATG	187						
QY	247	aagtggtgaggaagccctgattgactctggagaggaggtttcttgaaaggagccattcc	306						
DB	188	AAGCATGAGAGAGCCCTGATGATCTTGAGAGAGAGTTTCTGGAGAGCATTTTCC	247						
QY	307	cactgctgaaagagctaaacaggaatttgaaatctgtttccagcaatgaaagagatgaa	366						
DB	248	CACTGCTGMAAGAGCTTAACAGAGATTTGGAATTTGTTTCAGCAATGGAAGAAATGA	307						
QY	367	aggaaatcgcgctttccctccatgacgcgcggaatttggatggagggagagagacca	426						
DB	308	AGGAGATCCGGTCTTCCCTCAAGCCTCGGATTTGGATGGGAGAGAGAGACN	367						
QY	427	tggagagcctctctaaaggaagagccgctgcttggagagagcttgagaaacacagg	486						
DB	368	TTGAGGACCGGTGTCAAGAGGAGCCCGCTGCTTGTGGAGGAGTTGAGAAAACAAGG	427						
QY	487	cttcacctgtatcccaacttctcctgggctgctcctctgcaatgatctctcca	546						
DB	428	CTCAGCCTGTGATCCATTTCACTCGGGCTGTGCTCTCCATGATGATCTGCTCCA	487						
QY	547	tattttccgaagcgtttcattataaagatcagaatttcttaactgtagtggaaaat	606						
DB	488	TTATTTCCATAACGTTTTCATATAAAGATCAGCAATTTCTTAATTAATGGAAGAT	547						
QY	607	tgaatgaaacatcagatgtgaagcaccctccctggatccagatgatcaataatttccca	666						
DB	548	TGATGAAACATCATTTGATTTAGCAACCCCTGAAATCAGTATCGATATTTCTC	607						
QY	667	ctgattctattttcccggaagaccctaaagatcagatctaaacactgcctttttag	726						
DB	508	CTATCATTTGATTACTTCCCGGAACTACACAAATTACTTAAAACGTTCTTTTATGA	667						
QY	727	aagatgatatttggagaaagtgaagaagaccacccaagaatcgtatggacatcaaacaccttc	786						

```
568 AAGATGATATATTTTGGAAAAAGTAAAGAACACCAACCAAGCAATCAATGGACATGAACACCCCTC 727
Oy 787 ggaacttatgatctctctctctctctctctctctctctctctctctctctctctctctctctc 846
Db 728 AGAGCTTATGATGCTCTCTCTGATGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 787
Oy 847 aactctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 906
Db 788 ATTTACTATTTGAACTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 847
Oy 907 caaacaacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 966
Db 848 CGAAGGACCAACCTGCTGATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 907
Oy 967 taaagtcagagagagatgaactgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 1026
Db 908 CTAAAGTCCAGGAGAGATGACAGTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
Oy 1027 aagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1086
Db 968 ACAGAGGACCAATGCTTACACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027
Oy 1087 tcaatccacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1146
Db 1028 TTCTCCACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
Oy 1207 tcccaacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1266
Db 1148 TTCCCAACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1207
Oy 1267 aagatgaactctctctctctctctctctctctctctctctctctctctctctctctctctctct 1326
Db 1208 AAGTAAATATCTGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1267
Oy 1327 ccagatgagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1386
Db 1268 CGGCATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1327
Oy 1387 tgaatgagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1446
Db 1328 TGGTGAACCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1387
Oy 1447 cctctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 1478
Db 1388 CTTCTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1419
```

RESULT 15

```
US-08-201-118-6
: Sequence 6, Application 05/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: ROYCE A. ROYCE
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
: TITLE OF INVENTION: DNAs FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Kourile and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/201.118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/884.962
FILING DATE: 09-APR-1992
INVENTOR/AGENT INFORMATION:
NAME: ROYCE, ROYCE A.
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2422
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2009 base pairs
MOLECULE TYPE: cDNA
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-201-118-6
```

```
Query Match 72.4% Score 1264.6 DB 1: Length 2009;
Best Local Similarity 84.4% Pred. No. 0;
Matches 148; Conservative 0; Mismatches 254; Indels 22; Gaps 5;
Oy 1 cttcaatgagatcttctgtgctctgtctctctctctctctctctctctctctctctctctca 60
Db 195 CTTCAATGATCCACGCTGGCTCTGTGCTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 254
Oy 61 tctgagacagactctgtgagagagaaactctctctctctctctctctctctctctctctctca 120
Db 255 TCTGAGGACAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 314
Oy 121 tgaataatctctacagatagatgaagatctcaacaatctcaacaatctcaacaatctca 180
Db 315 TTGGAATATCTCGATGATGTAAGGACATGACAAATCTTTAACCAATTTCTCAA 374
Oy 181 aaactctatggcctgtgtctctctctctctctctctctctctctctctctctctctctct 240
Db 375 AAGTCTATGGCCCTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 434
Oy 241 gatatgagtgagagagagagagagagagagagagagagagagagagagagagagagagagag 300
Db 435 GATATGAACAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 494
Oy 301 attctctctctctctctctctctctctctctctctctctctctctctctctctctctctctc 360
Db 495 GTTTTCAGTGGCTGAAAAGGTAAACAAAGGACTTGGAAATCTTTTCAGCAATGGAAGA 554
Oy 361 gatgagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 420
Db 555 GATGAGAGAGATCCGGCTTCTGCTCTCTGACTCTGCGGAAATTTGGATGGGAGAG 614
Oy 421 gaaactatgagagagagagagagagagagagagagagagagagagagagagagagagagag 480
Db 615 GGAGCATCGAGGAGCTGTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 674
Oy 481 ccaaggctcctctctctctctctctctctctctctctctctctctctctctctctctctctct 540
Db 675 CCAATGCTCACCTGATCCACATTCCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 734
Oy 541 gctccatattctccagagagagagagagagagagagagagagagagagagagagagagagag 600
Db 735 GCTCTGTATTTTCCATGCTGATTTGATATAAAGATCAGAGGTTTCTTAAGTTGATGG 794
Oy 601 aaaaatgaaagaaacatcaagagatgaagcagagagagagagagagagagagagagagagagag 660
Db 795 AAAAAATCAATGAAGAACCTCAGGATCTGAGCTCTCCATCGATCGATCGATCGATCGATCGA 854
```

Db 1934 GTGTGAGTATGTTATCACTAGAGA 1958
Search completed: April 19, 2002, 08:10:30
Job time: 2930 sec

Oy 661 ttccacatactgatttctcggagaccatacaaatctttaaacccttctt 720
Db 855 TCCCTGCTCTCGATTATCCCGAAGTCATAAATAAGTGAATAATTTGCTT 914
Oy 721 ttatgaaagtatattttgagaaagaaacacacaaatcgtatgaacaaca 780
Db 915 ACATTAAAGATTATGATTGGAGAAATAAAGAAACATCAAGAAATCCCTGGACATGAACA 974
Oy 781 accctcgagcttatttattggtctctctcaaatggagaaagcaaaccaac 940
Db 975 GTCTCGGACCTTTATTTGATTTCTTGATCMAATGMAAGAAAGACACATCAAC 1034
Oy 841 agctcgaattcatttgaattcgttaactcactcagctgaacttacttgaactcgtgga 900
Db 1035 AGTCTGATTTACTTGTGAAGCTTGATGCCACTGTAAGTATGTTTGGGGCTGGAA 1094
Oy 901 cagagcaacaaagcaacacccctgagatgctctctctctcgtgagacacccagagg 960
Db 1095 CAGAGCAACGAGCAACCACTCTGAGTATGGACTCTGCTGCTGAGTATCCAGAGG 1154
Oy 961 tcaagctcaagtccaggaagattgaagctgtaattgacgaacacggagccctctga 1020
Db 1155 TCAACCTTAAGTCCAGAGAGATTTGATGTGAGTTGGCAGAACCGAGCCCTTGTA 1214
Oy 1021 tgaacagacagggccacatgcccacacagatgctgctgacagaggtccagagataca 1080
Db 1215 TGCAGGACAGGAGTCCACATGCCCTACACAGATGCTGTGTGCACGAGATCCAGAGATA 1274
Oy 1081 tgaacacatacccaacccagctgcccctcagtgacctgagctgacgtttaaattcagaacct 1140
Db 1275 TTGACCTCTCTCCCAACCACTGCCCCACCTGACCTGATGTTAAATTCAAAACCT 1334
Oy 1141 acctcattcccaagagcaacacatatgaactctcactctcactctcactctcactcaaca 1200
Db 1335 ACCTCATCCCAAGGGCCAGCACCANTATAGACTCTGACTCTGTGTGCACAAATGACA 1394
Oy 1201 aagaatttcccaacccagagatgctgacctcgtcactctctcgtgagaaagtggaaatt 1260
Db 1395 AAGAATTCGCCAACCCAGAGATGTTGACCTGGCCACTTCTGGATAAGAGTGGCACT 1454
Oy 1261 ttaagaaagtaactacttcatgcttctcagcaggaacacggatttggggggagg 1320
Db 1455 TTAGAAAGATGACTTCAATGCTTTCTAGCAGGAAACGGATGTATGGAGAGG 1514
Oy 1321 gctcgccacagatgagctgttttattctcgaactcttattcagaactttaacctga 1380
Db 1515 GCCTGCCCGCATGGAGCTGTTTATTCTGACCCACCATTTTGCAGAACTTTAACCTGA 1574
Oy 1381 aactctgattgacccaaaggaaccttgacacactcctgtgtcaatggattgtctctg 1440
Db 1575 AATCTCAGGTGGCCAAAGGATATTGACATACCCCAATGCCAATGCAATTTGGTCGTG 1634
Oy 1441 tccgacctctatcagctgactctcactctcactctcagagagcagagagctcagctg 1500
Db 1635 TCCACCCCTGTACACAGCTCTGCTTCATCTCTGTGAGAGGCGAGATAGTTGGCTG 1594
Oy 1501 ctctctgctgctc-cctcagctctctctctctctgcca-----aatctcact 1548
Db 1695 CTCTCTGCTGTGACCTGCAATTCCTGTTATCAGGGCCATTAGCTCTCTCTCTCTCT 1754
Oy 1549 atctgtgactctctctgacccgctcactcactcttccc- tcccccaagatctagt 1607
Db 1755 GTAGAGGATATTTCTGACTGTGATCCACATCTCCATTCCTCCAGATCCATG 1814
Oy 1608 aacttcagctccacttaaaa-----aattcactgcaaatatatactctattc 1659
Db 1815 AACATCAACCTCCATTAAAGAGATTCTTGGGTCACTTCTTAAATATACTCTATTC 1874
Oy 1660 cccctactataatagttacattgagtgccacataatgctgatactctgctaaagtga 1719
Db 1875 TCAATACTCTGTATCACTGTATTGAGCACCACAT-ATGCTAATACCTATCTACTCTGA 1933
Oy 1720 gttatcaacatattatttaata 1744

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 09:24:59 ; Search time 4158.4 Seconds
(without alignments)
2940.554 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcagaataatttgagcctg.....ttagctcatgtgaacgggg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 824589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vl.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.om.*
20: em.or.*
21: em.ov.*
22: em.pat.*
23: em.ph.*
24: em.pl.*
25: em.ro.*
26: em.sts.*
27: em.sy.*
28: em.un.*
29: em.vi.*
30: em.htgo_hum.*
31: em.htgo_inv.*
32: em.htgo_rod.*
33: em.htg_hum.*
34: em.htg_inv.*
35: em.htg_rod.*
36: em.htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	739.8	99.6	743	9	HUM2C9X02
2	728.8	98.1	143087	2	LI6878 Homo sapien
3	728.8	98.1	173154	2	AL1359672 Homo sapi
4	635.6	85.5	205791	2	AL133513 Homo sapi
5	433.6	58.4	143087	2	AL583836 Homo sapi
6	346.4	46.6	156492	2	AL359672 Homo sapi
7	331.2	44.6	654	9	AL157835 Homo sapi
8	325	43.9	173154	2	HUM2C18X02
9	325	43.9	173154	2	AL133513 Homo sapi
10	280	37.7	711	4	AL583836 Homo sapi
11	264.6	35.6	463	9	M74201 Rabbit cyto
12	238.6	32.1	147838	2	AF136833 Homo sapi
13	225.8	30.4	400	11	AC038888 Rattus no
14	221.8	29.9	168457	2	GI4580 Human STS S
15	217.2	29.2	513	10	AC078913 Mus muscu
16	178	24.0	163733	2	M18357 Rat cytochr
17	163	21.9	1473	22	AC083911 Rattus no
18	163	21.9	1473	22	E10631 Human CDNA
19	163	21.9	1576	9	E10853 CDNA encodi
20	163	21.9	1577	9	M15331 Human liver
21	163	21.9	1814	9	M21939 Human cytoc
22	163	21.9	1845	9	S46963 putative CY
23	163	21.9	1852	6	M61857 Human cytoc
24	163	21.9	1854	6	AR071579 Sequence
25	163	21.9	1854	9	AR071576 Sequence
26	161.4	21.7	1826	9	M61855 Human cytoc
27	161.4	21.7	1843	6	D00173 Homo sapien
28	153.4	20.6	1892	6	E02279 Human liver
29	152.4	20.5	1441	9	AR071581 Sequence
30	151	20.3	626	10	M21940 Human cytoc
31	149.6	20.1	1444	6	M33545 Rat female-
32	149.6	20.1	1473	22	E14930 Artificial
33	149.6	20.1	1669	6	E10866 CDNA encodi
34	149.6	20.1	1746	6	E14931 Human mRNA
35	149.6	20.1	1746	9	AR071575 Sequence
36	149.6	20.1	2395	9	M61854 Human cytoc
37	144	19.4	200	6	L07093 Human cytoc
38	144	19.4	200	22	HUMSVPA450A
39	140.2	18.9	1473	22	AR048195 Sequence
40	140.2	18.9	1995	9	E10095 Synthetic m
41	140.2	18.9	2009	6	E10865 CDNA encodi
42	140.2	18.9	2258	6	M61856 Human cytoc
43	140.2	18.9	2258	9	AR071577 Sequence
44	135.8	18.3	1473	22	AR071580 Sequence
45	135.8	18.3	1473	22	M61853 Human cytoc
					E10639 Human CDNA
					E10862 CDNA encodi

ALIGNMENTS

RESULT 1					
HUM2C9X02					
LOCUS	HUM2C9X02	743 bp	DNA	PRI	08-FEB-1999
DEFINITION	Homo sapiens cytochrome P450C9 (CYP2C9) gene, exons 2 and 3.				
ACCESSION	LI6878				
VERSION	LI6878.1	GI:291608			
KEYWORDS	CYP2C9; cytochrome P450; mephenytoin 4-hydroxylase.				
SEGMENT	2 of 7				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (sites)				
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	Goldstein,J.A., Raucy,J.L., Blaisdell,J.A., Faletto,M.B. and Romkes,M.				
JOURNAL	Cloning and expression of complementary DNAs for multiple members of the human cytochrome P4501C subfamily				
MEDLINE	Biochemistry 30, 3247-3255 (1991)				
REFERENCE	91182740				
	2 (bases 1 to 743)				

AUTHORS de Morais,S.M., Schweikl,H., Blaisdell,J. and Goldstein,J.A.
 TITLE Gene structure and upstream regulatory regions of human CYP2C9 and CYP2C18
 JOURNAL Biochem. Biophys. Res. Commun. 194 (1), 194-201 (1993)
 MEDLINE 93326116
 REFERENCE 3 (sites)
 AUTHORS Romkes,M., Faletto,M.B., Blaisdell,J.A., Raucy,J.L. and Goldstein,J.A.
 TITLE Correction: Cloning and expression of complementary cDNAs for multiple members of the human cytochrome P4501C subfamily
 JOURNAL Biochemistry 32, 1390-1390 (1993)
 MEDLINE 93192243

FEATURES
 source Location/Qualifiers
 1. .743
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /tissue_type="liver"
 /tissue_lib="EMBL3 library from J.A. Goldstein"
 131..293
 /gene="CYP2C9"
 /citation=[1]
 /evidence=experimental
 /number=2
 465..614
 /gene="CYP2C9"
 /citation=[1]
 /citation=[3]
 /number=3
 /evidence=experimental
 BASE COUNT 182 a 152 c 196 g 213 t
 ORIGIN

Query Match 99.6%; Score 739.8; DB 9; Length 743;
 Best Local Similarity 99.7%; Pred. No. P.le-197;
 Match's 741; Conservative 0; Mismatches 2; Indels 0; Gaps 0:

OY 1 tcagaaatttgagcctgtgtgctgaataaaagcgtacacaaatacaataatca 60
 |||||
 DB 1 TCAGAAATATTTGAAGCCTGTGTGCTGAATAAAGCATACAAATACAAATATCA 60
 |||||
 OY 61 tgcataatcaggcttagcaaaatggaacaaatagtaacttcgttcttattctgtct 120
 |||||
 DB 61 TGTAAATCAGGCTTAGCAATAGGACAAATAGTAACTTCGTTTCGTGTATCTCTGCT 120
 |||||
 OY 121 actttcctagctctcaaaagtctatggccctgtgttcactctgtatttggcctaaaacc 180
 |||||
 DB 121 ACTTCTTAGCTCTCAAGGCTCTATGGCCCTGTGTCTACTCTGTATTTGGCCTCAAAACC 180
 |||||
 OY 181 catagtgtgctgcattgatatgaagcagtggaagcctcctgattgattgagaggaga 240
 |||||
 DB 181 CATAGTGGTGTGCATGGATATGAAGCAGTGAAGGAAGCCCTGATTGATCTTGGAGAGGA 240
 |||||
 OY 241 gttttctggaagaggcattttccctcctggtgctgaagagcgtacacagagatttggtagtg 300
 |||||
 DB 241 GTTTCCTGGAGAGGCATTTCCCACTGGCTCAAGAGCTAACAGAGATTGGTAGGTG 300
 |||||
 OY 301 tgcattgctgttttcagcatctgtctgggagtgaggagtggaacagagacttac 360
 |||||
 DB 301 TGCATGTGCTGTTTCAGCATCTGTCTGGGGATGGGAGGATGGAACACAGAGACTTAC 360
 |||||
 OY 361 aagagctcctcgagcagagcttggcccatccacatggctcctccagtgctcagcttctctt 420
 |||||
 DB 361 AGAGCTCCTCGGCAGAGCTTGGCCCATCCACATGGCTGCCAGTGTCCAGCTTCTCTT 420
 |||||
 OY 421 cttgctgggattcctcctcctagtttcgtttctctctcctccttagaattgttttcagcaa 480
 |||||
 DB 421 CTTGCTGGGATCTCCTCCTAGTTTCGTTTCTTCTTCCTGTTAGGAATGTTTTCAGCAA 480
 |||||
 OY 481 tgaagaagaatgaagagatccggcgtttctccctcctcctcagctcgcgaattttgggat 540
 |||||
 DB 481 TGAAGAAGAATGGAAGAGATCCGGCGTTTCTCCCTCAATGACGCTCCGGAATTTGGGAT 540
 |||||

OY 541 ggggaagaggagcattgagaccggtgttcaagagaagcccgctccttctgtgagagatt 600
 |||||
 DB 541 GGGGAAGAGGAGCATTGAGGACCGGTGTTCAAGAGGAAGCCCGCTCTGTGTGGAGAGTT 600
 |||||
 OY 601 gagaataaaccaagggtgggtgacctactccatatacactgaccttactggagactactatct 660
 |||||
 DB 601 GAGAAAAACCAAGGTTGGGTGACCTTACTCCATATACATGACCTTACTGGACTACTATCT 660
 |||||
 OY 661 tctctactgacattcttggaacatttcagggtgagccatatacttttcattatgagctcgg 720
 |||||
 DB 661 TCTCTACTGACATTTCTTGAAACATTTTCAGGGTGGCCATATCTTTCATTATGAGTCTGG 720
 |||||
 OY 721 ttgttagctcatgtgaacgggg 743
 |||||
 DB 721 TTGTTAGCTCATGTGAAGCGGG 743
 |||||

RESULT 2
 AL359672 143087 bp DNA HTG 27-JUN-2001
 LOCUS Homo sapiens chromosome 10 clone RP11-208C17, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 unordered pieces.
 AL359672
 ACCESSION AL359672
 VERSION GI:14575223
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Johnson,C.
 Direct Submission
 Submitted (26-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Jun 28, 2001 this sequence version replaced gi:14529836.
 ----- Genome Center
 Center: Sanger Centre
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information
 Center project name: BA208C17
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; L08752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 142666 bases at least Q40
 Consensus quality: 142843 bases at least Q30
 Consensus quality: 142920 bases at least Q20
 Insert size: 142987; sum-of-contigs
 Insert size: 136404; agarose-fp
 Quality coverage: 8.55x in Q20 bases; sum-of-contigs Quality
 coverage: 7.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 consists of 2 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

* 1 47005: contig of 47005 bp in length
 * 47006 47105: gap of 100 bp
 * 47106 143087: contig of 95982 bp in length.

FEATURES
 source

Location/Qualifiers
 1. .143087
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="*10"
 /clone="RP11-208C17"
 /clone_lib="RPC1-11.1"

```
misc_feature 1.. 47005
/notes="assembly_fragment:00494
clone_end:SP6
vector_side:left"
misc_feature 47106..1143087
/notes="assembly_fragment:00967"
BASE COUNT 41247 a 26155 c 27788 g 47797 t 100 others
ORIGIN

Query Match 98.1%; Score 728.8; DB 2; Length 143087;
Best Local Similarity 99.6%; Pred. No. 1.5e-194;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcatcaacaatacaatacaatca 60
Db 60120 TCAGAAATATTGAAGCCTGTGTGCTGAATAAAGCATACAAATACATAATCA 50179
Qy 61 tgcataatcagccttagcaaatgacaaatagtaacttcgtttgctgttatctctct 120
Db 60180 TGCTAAATCAGGCCTAGCAAAATGACAAATAGTAACCTCGTTTGCTGTA 50239
Qy 121 acttcttagctctcaaggtctatgcccctgtgttcactctgtattttggcctga 180
Db 60240 ACTTCTTAGCTCTCAAGGCTATGGCCCTGTGTCACICTGTATTTGGCCCT 50299
Qy 181 cataagtgctcatagatgaagcagtgaaagagccctgattatctctgagaga 240
Db 60300 CATAGTGGTCTGATGATGAAGCATGAAGCATGAAGAGCCCTGATGATCT 50359
Qy 241 gtttcttgaagagcattttcccaactggctgaaagagctaaacagagatttga 300
Db 60360 GTTTCTGGAAGAGGCATTTTCCCACTGGCTGAAAGAGCTAACAGAGATT 50419
Qy 301 tgcattgaccttttcagacatctcttgggatggagatggagatgaaacagact 360
Db 60420 TGCAATGCTCTTTCAGCATCTGCTTGGGATGGGAGGATGAAACAGACAT 50479
Qy 361 agagctctcgggcagagcttggccatccacatggctgcccagtgctcagctct 420
Db 60480 AGAGCTCTCGGCAGAGCTTGGCCCATCCACATGGCTGCCAGTGTGAGT 50539
Qy 421 ctgctcagggatcctccttagtttgccttctctctctctctctctctctctca 480
Db 60540 CTGCTCGGATCTCCCTCCTAGTTTCTCTCTCTCTCTCTCTCTCTCTCT 50599
Qy 481 tggaaagaaatggaagagatccgcgctttctccctcatgacgctgcggaatt 540
Db 60600 TGGAAAGAAATGGAAGAGATCCGCGCTTCTCCCTCATGACGCTGCGGA 50559
Qy 541 ggggaagaggagcattgaggaccgtgttcagagagaaagccgcctgtgagagg 500
Db 60660 GGGGAAGAGGAGCATTGAGGACCGGTGTTCAAGAGAGGAAGCCGCTGCT 50719
Qy 601 gagaataaccaggtggtgacctactccatcatcactgacctactgagactact 560
Db 60720 GAGAAACCAGGTTGGGTGACCCCTACTCTATCATATCATTGACCTACTAT 50779
Qy 661 tctctactgacattcttggaaacatttcagggggtggccatatcttccattatga 719
Db 60780 TCTCTACTGACATCTTGGAAACATTTTCAGGGGIGGCCATATCTTTCATTAT 50839
Qy 720 gtttgaactcattgaaagcagg 743
Db 60840 GTTGTAGCTCATGTGAAGCGGGG 60863

RESULT 3
AL133513
LOCUS Homo sapiens chromosome 10 clone RP11-400G3, 04-JUL-2001
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AL133513
```

```
AL133513.11 G1:14575067
HTG: HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
human.
ORGANISM Homo sapiens
REFERENCE Brown,J.
AUTHORS Direct Submission
TITLE Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 18A, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT requests: clonerequest@sanger.ac.uk
On Jun 28, 2001 this sequence version replaced gi:14456168.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA400G3
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 172989 bases at least Q40
Consensus quality: 173136 bases at least Q30
Consensus quality: 173145 bases at least Q20
Insert size: 173154; sum-of-contigs
Insert size: 116005; 33.1% error; agarose-fp
Quality coverage: 6.48x in Q20 bases; sum-of-contigs Quality
coverage: 9.80x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..173154
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-400G3"
/clone_lib="RPC1-11.2"
/misc_feature 1..173154
/notes="assembly_fragment:00748
clone_end:SP6
vector_side:right"
BASE COUNT 49163 a 32272 c 33707 g 58012 t
ORIGIN

Query Match 98.1%; Score 728.8; DB 2; Length 173154;
Best Local Similarity 99.6%; Pred. No. 1.5e-194;
Matches 741; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcatcaacaatacaatacaatca 60
Db 150754 TCAGAAATATTGAAGCCTGTGTGCTGAATAAAGCATACAAATACATAATCA 150813
Qy 61 tgcataatcagccttagcaaatgacaaatagtaacttcgtttgctgttatctctct 120
Db 150814 TGCTAAATCAGGCCTTAGCAAAATGACAAATAGTAACCTCGTTTGCTGTA 150873
Qy 121 acttcttagctctcaaggtctatgcccctgtgttcactctgtattttggcctga 180
Db 150874 ACTTCTTAGCTCTCAAGGCTATGGCCCTGTGTCACICTGTATTTGGCCCT 150933
Qy 181 cataagtgctcatagatgaagcagtgaaagagccctgattatctctgagaga 240
Db 150934 CATAGTGGTCTGATGATGAAGCATGAAGCATGAAGAGCCCTGATGATCTT 150993
Qy 241 gtttcttgaagagcattttcccaactggctgaaagagctaaacagagatttga 300
```

```
Db 150994 GTTTCTGGAAGAGGCAATTTCCCACTGGCTGAAAGAGCTAACAGAGGANTTTGGTAGGTG 151053
Qy 301 tgcattgctctgttccagcatctgtcttgggagatggggagatggaacacagagacttac 360
|||||
Db 151054 TGCATGTGCTCTTTTCCAGCATCTGTTGGGATGGGAGGATGGAACACAGAGACTTAC 151113
|||||
Qy 361 agagctctctggcagagattggccatccacatggctgcccagtgctcagctctctcttt 420
|||||
Db 151114 AGAGCTCTCTGGGACAGAGTTGGCCATCCACATGGCTGCCAGGTGTCCAGCTTCCTCTTT 151173
|||||
Qy 421 ctgtcctgggagatccctcctcagttctgtctctctcctgttagaattgtttcagcaa 480
|||||
Db 151174 CTGGCTGGGATCTCCCTCTAGTTCTGTTCTCTTCTGTTAGGAATTTGTTTCAGCAA 151233
|||||
Qy 481 tgaagaagaaatgaagagagatccggcggtttctcctcctcctcagcctcggaatttgggat 540
|||||
Db 151234 TGAAGAAGAAATGGAAGGAGATCGGGGTTTCTCCCTCATGACGCTCGGAAATTTGGGAT 151293
|||||
Qy 541 gggagaagagagatgagacagctgttcaagaggaagccgctgcttctgtggaagagt 600
|||||
Db 151294 GGGGAAGAGGAGCATTGAGACCGTGTTCAGAGGAAGCCGCTGCTTGTGGAGGAT 151353
|||||
Qy 601 gagaaaaaaccaaggtgggtgacccctactccatcatcactgacacttactggactactatct 660
|||||
Db 151354 GAGAAAAACCAAGGTTGGGTGACCTTACTCCATATCAGTACGCTTACGAGCTACTACT 151413
|||||
Qy 661 tctctactgacattctggaacatttcaaggtgacatctcttctcattatgaagt-ctg 719
|||||
Db 151414 TCTCTACTGACATCTTTGGAACATTTTCAGGGGTGGCCATCTTTTCATTAATGAGTCCCTG 151473
|||||
Qy 720 gttgttagctcatgtgaagcgggg 743
|||||
Db 151474 GTTGTAGTCATGTGAAGCGGGG 151497
|||||
```

RESULT 4

```
AL583836 205791 bp DNA HTG 13-JUN-2001
LOCUS Homo sapiens chromosome 10 clone RP11-466J14, *** SEQUENCING IN
DEFINITION PROGRESS ***, 2 unordered pieces.
ACCESSION AL583836
VERSION AL583836.11 GI:14455940
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 205791)
Mashreghi-Mohammadi,M.
Direct Submission
Submitted (11-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
On Jun 14, 2001 this sequence version replaced gi:14148873.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA466J14
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 205236 bases at least Q40
Consensus quality: 205448 bases at least Q30
Consensus quality: 205560 bases at least Q20
Insert size: 205691; sum-of-ctnigs
Insert size: 188405; 9.1% error; agarose-fp
Quality coverage: 11.31x in Q20 bases; sum-of-ctnigs Quality
coverage: 12.57x in Q20 bases; agarose-fp
-----
```

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved
*
* 1 88240: contig of 88240 bp in length
* 88241 88340: gap of 100 bp
* 88341 205791: contig of 117451 bp in length.
* Location/Qualifiers
* 1..205791
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="10"
* /clone="RP11-466J14"
* /clone_lib="RPC1-11.2"
* 1..88240
* /note="assembly_fragment:00191"
* fragment_chain:1"
* 88341..205791
* /note="assembly_fragment:01703"
* fragment_chain:1"
* clone_end:17
* vector_side:right"
* BASE COUNT 56694 a 38262 c 40890 g 69845 t 100 others
* ORIGIN
```

```
Query Match 85.5%; Score 635.6; DB 2; Length 205791;
Best Local Similarity 93.5%; Pred. No. 2.8e+168;
Matches 695; Conservative 0; Mismatches 44; Indels 4; Gaps 3;
```

```
Qy 1 tcagaaatattgaagcctgtgtgctgaataaaagcaccatacaatacaataatca 60
|||||
Db 96936 ICAGAAATATTG-AGCCTGTGCTGAATAATAAGCATAACAATAAATAATGA 96994
|||||
Qy 61 tctcaaatcagcttagcaaaatggacaaatagtaactctgttctcttactctgct 120
|||||
Db 96995 ATCTAAGTCAGGCTTAGTAATATGGACAAACAGTACATTCATTTGCTTAACTGTA 97054
|||||
Qy 121 acttctcagctctcaaggtctatggcctgtgttcactctgtattttggcctgaaacc 180
|||||
Db 97055 CCTTTTCTAGCTCTCAAAAATCTATGGCCCTGTGTTCTACTCTGTATTTTGGCCTGGAACG 97114
|||||
Qy 181 catagtgtctcctcagatgagatgagcagtggaagagccctgattgcttgaggaga 240
|||||
Db 97115 CATGTGGTGTCTGTCATGGATATGAAGTGGTGAAGGAAGCCCTGATTTGAGAGGA 97174
|||||
Qy 241 gttttctgaaagagcattttccctcactggctgaaagagctaacagaggtttggtagg 300
|||||
Db 97175 GTTTCCTGGAAGAGGCCATTTCCCACTGCTGAAAGAGCTAACAGAGGATTTGGTAGGTG 97234
|||||
Qy 301 tgcattgctcttccagcatctgtcttgggagatggggaggtggaacacagagacttac 360
|||||
Db 97235 TGCAAGTGGCTGTTTTCAGCATCTGCTTTGGGATGGGGAGGATGGAAAAC--AGACTAGC 97292
|||||
Qy 361 aagctctctcggcagagcttggcccatccacatggctgcccagtgctcagctctctctt 420
|||||
Db 97293 AGAGCTTCTCGGCGACAGCTTGGCCCATCCACATGGCTGCCAGTGTCACTCTCTT 97352
|||||
Qy 421 ctgtcctggagatccctcctcagttctgttctcctcctcctcctcctcctcctcctc 480
|||||
Db 97353 CTTGGCTGGGATCTCCCTCTCCTAGTTTCGTTTCTCTTCTCTTCTAGGAATCGTTTTCAGCAA 97412
|||||
Qy 481 tgaagaagaaatgaagagagatccggcggtttctcctcctcagcctcggaatttgggat 540
|||||
Db 97413 TGAAGAGAGATGGAAGAGAGATCGGGCTTTCTCCCTCATGACCTCGCGAAATTTGGGAT 97472
|||||
Qy 541 gggagaagagagcattgagagacagctgttcaagaggaagcccgctgctgtgtgagagatt 600
|||||
Db 97473 GGGGAAGAGGAGCATTCAGGACCGTGTTCAGAGGAAGCCCGCTGCTTGTGGAGGAGTT 97532
|||||
```



```
Db 419 CAGGCTTCATTCGGCATCTGCTGTCAGCCC 451
|| |||| | | ||||| ||| |
AC083888 147838 bp DNA HTG 01-APR-2001
Rattus norvegicus clone RP32-307N15, WORKING DRAFT SEQUENCE, 21
unordered pieces.
AC083888
AC083888 12 GI:13493083
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Norway rat
ORGANISM
Rattus norvegicus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
(bases 1 to 147838)
AUTHORS
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S., L., Anaratunga, H.C., Are, J.R., Banks, T., Barbarella, J.,
Benton, J., Binage, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, I.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Forster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, I.,
Gasta, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C.,
Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, Z., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozada, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metker, M., Miner, G., Miner, Z., Mitchell, J., Mohabbat, K.,
Morgan, A., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
Ogulu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, F., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 147838)
Worley, K.C.
Direct Submission
Submitted (05-OCT-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Apr 1, 2001 this sequence version replaced gi:13253418.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: IUAF
FEATURES
Source
Location/Qualifiers
1..147838
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="RP32-307N15"
BASE COUNT 4949 a 29290 c 29368 g 42091 t 2140 others
ORIGIN
```



```
Db 27819 TTTGCTAAATAAAAAATTAATAACAGACATAATATCAGCATGAATGAGGTTACACT 27878
Qy 81 atgagcaaaatagtaac-ttcgtttgtgttatctctgtctacttctct-----agctc 133
Db 27879 ATATTGATCAAGGTTACATAATGTTGCTAATAATTTGCAATTTCTTATTTCCAGTTC 27938
Qy 134 tcaaaaggtctatggccctgtgttcaactctgtattttggcctgaacccatagtggtctg 193
Db 27939 TCAAAAGCTTATGGCCGTGTCTCACTCTGTATTGGGCTCAAAAGCCTACTGTGATATTG 27998
Qy 194 catgatataagcagcagtgaaagaaacccctgtattgatctctggagagagatttcttggaga 253
Db 27999 CATGGTATGAAGCAGTGAAGGAAGCTCTGATTGATCGTGGGAGGAGTTTGTCTGGTAGA 28058
Qy 254 ggcattttccactggctgaaagagctaacagagagatttggtaggtgtgcattgtgcctgt 313
Db 28059 GGAAGCTTTCCAAATGGCTGCAAAAATTTATTAAGGTTTGTGTAAGT----ATGTGATAC 28114
Qy 314 ttcagcatctgtcttgggagtggaagagtggaacacagaaacttacagagcctctcagg 373
Db 28115 ATTATGTGTGCTTCACAGTGTTTACAGTGAGATATAAGATTGAAAGAAAGGATGGG 28174
Qy 374 cagagcttgagccatccacatagctgcccagtgctcagcttctcttcttctgctgggagtc 433
Db 28175 CTTCCTTCGTAGGCAATGGGCCACTCTATGTTTCCACACTCACCTGTCTTAGCAT 28234
Qy 434 tccctcctagtttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 493
Db 28235 CTATTCTAGTACC--TTCTATTCTTTAGGCGTGTGTTTGTAGCAATGGAACAGATGG 28292
Qy 494 aaggagatccggcgttcttctcctcatgacgctgcggaattttggagatgggaagagagac 553
Db 28293 AAAGATATGAGAGATTTACACTCATGACCCCTGCGAAATTTGGGATGGGAAAGGAAC 28352
Qy 554 attgagaccgtgttcaagaggaagccgctgcttctgtgagagtggaagaaacccaag 613
Db 28353 ATTGAGACCGTGTTCAGAGGAAGACACAGTGTCTTGTGGAAGAACTGAGGAAACCAAA 28412
Qy 614 ggt 616
Db 28413 GGT 28415

RESULT 15
RATCY45M2
LOCUS Rat cytochrome P-450(M-1) gene, exons 2 and 3. 27-APR-1993
DEFINITION M18357
ACCESSION M18357.1 GI:203687
VERSION cytochrome P450.
KEYWORDS 2 of 8
SEGMENT Rat (male) liver DNA, (library of L.Jagodzinsky), clones
SOURCE lambda-(1,3,18,22,23,29).
ORGANISM Rattus
REFERENCE 1 (bases 1 to 513)
AUTHORS Morishima,N., Yoshioka,H., Higashi,Y., Sogawa,K. and Fujii-Kuriyama,Y.
TITLE Gene structure of cytochrome P-450(M-1) specifically expressed in male rat liver
JOURNAL Biochemistry 26, 8279-8285 (1987)
MEDLINE 88163490
FEATURES Location/Qualifiers
source 1..513
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
<1..11
intron /note="P450 intron A"
prim_transcript <1..>513
/note="P450 mRNA and introns"
exon 12..117
```

```
/number=2
175..318
/note="P450 intron B"
319..468
/number=3
469..513
/note="P450 intron C"
119 a 95 c 144 g 154 t
BASE COUNT About 1.7 kb after segment 1.
ORIGIN
```

```
Query Match 29.2%; Score 217.2; DB 10; Length 513;
Best Local Similarity 69.9%; Pred No 2.4e-50;
Matches 348; Conservative 0; Mismatches 123; Indels 27; Gaps 3;

Qy 121 acttctcagctctcaaaaggtctatggccctgtgttcaactctgtattttggcctgaaacc 180
Db 2 AATTTTCCAGTTTTCAAAAGCTTACGGCCCTATATTTACTCTGTATTTTGGGATGAAGCC 61
Qy 181 catagtgtgctcagatgatataagcagtgaaagagccctgattgatcttggagaga 240
Db 62 CTTTGTGGTGTTCATGGGTATGAAGCTGTGAAGGAAGCTCTTGTGTATCTAGGAGAGA 121
Qy 241 gttctctggaagagcattttcccaactggctgaaagagctaacacagagagatttggtaggtg 300
Db 122 ALTTCTGGAAGAGCGAGTTTCCAGTATCTGAAAGAGTTTAAACAAGGCCCTTGGTAAGTG 181
Qy 301 tgcaltgctctgttctcagcatctgtcttggggatgggagagatggaaaacagagacttac 360
Db 182 TGAGTACAAATGTAATTTCTGTGTGTGTGAGCCCTCGGCCCTTGGCAGA---TGAGACTTAA 237
Qy 361 agagctctcgggcagagcttgcccacatccacatggctgccagtgctcagcttctctcttt 420
Db 238 TGAGCTCAGAGCCCATTAGGATAGC---ACCATGTGTGTCAGTTCGTGCTGCTCCTTGACT 294
Qy 421 ctgctctgggagatctcctcctcctagtttcttcttcttcttcttcttcttcttcttctt 480
Db 295 AGT-----TTCCTGTTCCCAATTCCTGAGGAGTCAITTTTAGCAA 334
Qy 481 tggaaagaaatggagagatccagcgtttctcctctcatgacgctgcggaattttgggat 540
Db 335 TGGGATGCAATGGAAGGAGATCCGGCGTTCCTCCATCATGACCCTGAGGACTTTTGGGAT 394
Qy 541 ggggaagagagcatttaggaccctgttcaagagagagcccgctgccttcttggagagatt 600
Db 395 GGGCAAGAGGACCATTGAGGACCGTATTCAAGAGGAGGCTCAGTGCCTTGTGGAGGAAC 454
Qy 601 gaaaaaaaccagagtgag 618
Db 455 GAGGAAGAGCAAAAGGTAG 472
```

Search completed: April 19, 2002, 09:32:08
Job time: 7828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:08:00 ; Search time 232.27 Seconds
(without alignments)
885.492 Million cell updates/sec

Title: US-09-763-292-1
Perfect score: 1746
Sequence: 1 cttcaatggtccctttgtg.....acatattattattaaataga 1746

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 78474 seqs, 58898373 residues

Total number of hits satisfying chosen parameters: 156948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents_NA_New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	139.6	8.0	526	6	US-10-106-698-2440
2	137.2	7.9	591	6	US-10-106-698-858
3	59.2	3.4	465	6	US-10-103-313-627
4	51.8	3.0	1186	6	US-10-103-313-621
5	51.8	3.0	1186	6	US-10-105-299-7204
6	51.6	3.0	1454	5	US-09-559-013D-11
7	51.6	3.0	1464	5	US-09-559-013-11
8	49	2.8	2026	6	US-10-106-698-368
9	49	2.8	13327	6	US-10-105-299-7201
10	45.8	2.6	1530	1	PCT-US02-06912-12
11	45.8	2.6	1693	1	PCT-US02-06912-8
12	44.6	2.6	254	5	US-09-975-254-18019
13	43.6	2.5	832	6	US-10-103-313-626
14	43.6	2.5	1001	6	US-10-103-313-625
15	41.4	2.4	272	5	US-09-975-254-16494
16	41.4	2.4	1182	6	US-10-103-313-622
17	41.4	2.4	1182	6	US-10-105-299-7203
18	41	2.3	583	6	US-10-102-806-124
19	41	2.3	602	6	US-10-103-313-66
20	41	2.3	1745	1	PCT-US02-06912-6
21	40.8	2.3	252	5	US-09-975-254-11513
22	40.8	2.3	252	5	US-09-975-254-21714
23	40	2.3	248	5	US-09-975-254-23845
24	39.4	2.3	237	5	US-09-975-254-14003
25	39.4	2.3	2356	6	US-10-103-313-103
26	38.4	2.2	266	5	US-09-975-254-20507

27	38	2.2	1756	6	US-10-104-706-1
28	37.5	2.2	253	5	US-09-975-254-13380
29	37.2	2.1	261	5	US-09-975-254-19229
30	36.6	2.1	286	5	US-09-975-254-23331
31	36.2	2.1	267	5	US-09-975-254-19290
32	35.8	2.1	745	6	US-10-103-313-102
33	34.2	2.0	191	5	US-09-975-254-25520
34	34.2	2.0	7932	6	US-10-105-299-7189
35	33.8	1.9	1658	6	US-10-026-666-1
36	33	1.9	2650	1	PCT-US02-08992-16
37	32.8	1.9	257	5	US-09-975-254-12971
38	32.8	1.9	1614	1	PCT-US02-06912-4
39	32.4	1.9	3220	6	US-10-103-313-100
40	32.2	1.8	265	5	US-09-975-254-24906
41	32	1.8	1084	6	US-10-103-313-202
42	32	1.8	5115	5	US-09-630-630B-25
43	31.8	1.8	4540	6	US-10-105-299-9415
44	31.6	1.8	358	6	US-10-106-698-3502
45	31.6	1.8	748	6	US-10-106-698-1347

ALIGNMENTS

RESULT 1
US-10-105-698-2440
: Sequence 2440, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypep
: FILE REFERENCE: PA005P1
: CURRENT APPLICATION NUMBER: US/10/106,698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564
: SOFTWARE: PatentIn Ver. 3.0
: SEQ ID NO 2440
: LENGTH: 526
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (4)..(4)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (23)..(23)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (32)..(33)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (310)..(310)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (332)..(332)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (336)..(336)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (369)..(369)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (375)..(375)
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: misc_feature
: LOCATION: (387)..(387)
: OTHER INFORMATION: n equals a,t,g, or c

```

: NAME/KEY: misc_feature
: LOCATION: (401)..(401)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (414)..(414)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (419)..(419)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (461)..(462)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (488)..(488)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (518)..(518)
: OTHER INFORMATION: n equals a.t.g, or c
: NAME/KEY: misc_feature
: LOCATION: (523)..(523)
: OTHER INFORMATION: n equals a.t.g, or c
US-10-106-698-2440

```

Query Match 8.0% Score 139.6; DB 6; Length 526;

Best Local Similarity 57.5% Pred. No. 5.9e-27;

Mismatches 238; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```

Oy 1021 tgcacgagggccacatgccttacacagatgctgtgtgacaggggtccagagataca 1080
Db 15 ttcatacagagccaaannccatcacagagggcagtcacatctatgagattcagagattt 74

Oy 1081 tgaacctatccacacagcctgcccacgacgtgacgtgacgttgaattcagaact 1140
Db 75 ccgaccttcccatgggtgtgcccacattgtccacacacacacagcttccgaggt 134

Oy 1141 acctattcccaaggccacacacattaaacttccctcactctctgtgtacatgacacaa 1200
Db 135 acatcatcccaaggacacagaagtattctcctcctgagcactgctcctcagaccac 194

Oy 1201 aagaattcccaaccacagagatgttgacctctgaccttctgtgagagggggaatt 1250
Db 195 actacttgaaaccacagaccccttcactcctgaccttctgtgagagggggcag 254

Oy 1261 ttaagaaagttaactactctccttctcagcagagaaacggattgtgtggagagg 1320
Db 255 tgaagaaactgaagcttttaccctctctctttaggagagcgattgtgtggnagag 314

Oy 1321 gctgcccgcagatgagctgttttattcctgaccttcttaccagaaatttaacctga 1380
Db 315 gcatgcccgtgcggaantgtntcctctctcaccacacacacacacacacacacacac 374

Oy 1381 aatctctgattgaccccaaggaccttgacacacactctgtgtcaatggatttg 1434
Db 375 ncagcccgtgncggcagaagacatcmtcagacacccnagaangtggttg 428

```

RESULT 2

```

US-10-106-698-858
: Sequence 858, Application US/10106698
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
: FILE REFERENCE: PA00591
: CURRENT APPLICATION NUMBER: US/10/106,698
: PRIOR FILING DATE: 2002-03-27
: PRIOR APPLICATION NUMBER: PCT/US00/26524
: PRIOR FILING DATE: 2000-09-28
: PRIOR APPLICATION NUMBER: US 60/157,137
: PRIOR FILING DATE: 1999-09-29
: PRIOR APPLICATION NUMBER: US 60/163,280
: PRIOR FILING DATE: 1999-11-03
: NUMBER OF SEQ ID NOS: 8564

```

: SOFTWARE: PatentIn Ver. 3.0

: SEQ ID NO 858

: LENGTH: 591

: TYPE: DNA

: ORGANISM: Homo sapiens

: FEATURE:

: NAME/KEY: misc_feature

: LOCATION: (132)..(132)

: OTHER INFORMATION: n equals a.t.g, or c

: NAME/KEY: misc_feature

: LOCATION: (542)..(542)

: OTHER INFORMATION: n equals a.t.g, or c

: NAME/KEY: misc_feature

: LOCATION: (550)..(550)

: OTHER INFORMATION: n equals a.t.g, or c

: NAME/KEY: misc_feature

: LOCATION: (579)..(579)

: OTHER INFORMATION: n equals a.t.g, or c

: NAME/KEY: misc_feature

: LOCATION: (590)..(590)

: OTHER INFORMATION: n equals a.t.g, or c

: US-10-106-698-858

Query Match 7.9% Score 137.2; DB 6; Length 591;

Best Local Similarity 58.5% Pred. No. 2.6e-26;

Mismatches 238; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

```

Oy 972 gtccaggaagagattgaacgtgtcatctgttcacagaaaccgagccctgcatcacagagg 1031
Db 7 gtccacagagagatcgacgacgtatagaggcaggtgctgagcaccagagatgggtgaccag 66

Oy 1032 gcccacatccctacacagatgctgtgtgacagaggtgccagagatcacatcgacctc 1091
Db 67 gctcacatccctacacacactccctgattcatgaggtgacgctgttggggacatcgtc 126

Oy 1092 cccaccagcctgcccacatgacgtgacctgtgacgttaattcagaactacctattccc 1151
Db 127 cctgnggttgaccatcatgacatccctgacatcgaagtacaggggtccgcatccct 186

Oy 1152 aagggcacacacatttaactctcctcactctgtgtacatgacacaaagaattccc 1211
Db 187 aagggaacacacatcatcaccacactcactcgtgctgagagatgagggcgtctggag 246

Oy 1212 aaccagagatgttgacctcctcgtcacttctggtgagaggttgaaatttaagaaagt 1271
Db 247 aagcctctcgtctccaccccgacacactctcgtgagccagggccacttctggaagccg 306

Oy 1272 aactactctatgcttctcagcaggaacacgattgtgtggagagggcctggtccgcg 1331
Db 307 gaggcctctcgtccttctcagcagggccgctgcatgctcgggagccttggccgcg 366

Oy 1332 atggagctgttttattcctgacctcttctacagaaactttacagaaacttaacct 1378
Db 367 atgagctctctctcttctcactcctgctgacgacttcagctt 413

```

RESULT 3

```

US-10-103-313-627
: Sequence 627, Application US/10103313
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PJ207C1
: CURRENT APPLICATION NUMBER: US/10/103,313
: CURRENT FILING DATE: 2002-03-12
: NUMBER OF SEQ ID NOS: 653
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 627
: LENGTH: 465
: TYPE: DNA
: ORGANISM: Homo sapiens

```



```
; LENGTH: 2026
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1326)..(1326)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-368

Query Match      2.8%; Score 49; DB 6; Length 2026;
Best Local Similarity 43.0%; Pred. No. 0.002;
Matches 238; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Qy 849 ttcaattgaaacctgttaactcactgagctgaacttaacttaactgagctggaacagagaa 908
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 862 ttgactgtggatgagattgtggccaggccttcaattcttcttcttcttcttcttcttctt 803

Qy 909 acaagcacaacctgagatatactctctctctctctctctctctctctctctctctctct 968
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 802 atcacacacacactttcttttgcacacttacttacttgcgcacacacacacacacacacag 743

Qy 969 aaagtcaggaagagatgaacgtgtcattgagcagaacacacacacacacacacacacacac 1028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 742 aaagttctgagagagtagacgttttttaaggagaaacacacacacacacacacacacacac 683

Qy 1029 aggggccacatccctacacagatgctgtgtgacagaggtccagagatatacatgacctc 1088
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 ctgcaggaaggcctgccttcttctgacatggtgattgacagacacacacacacacacacac 623

Qy 1089 atccccacacagctgccccatgacgtgacgttgcgttgcgttgcgttgcgttgcgttgc 1148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 ccagcttttcagattcacacggcagcagctcagagactgcgaggttcggcgcagcgcacac 563

Qy 1149 ccaaggggcaacacattatcaacttccctcacttctgtgtgtgtgtgtgtgtgtgtgtgt 1208
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 562 cccgagggcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 503

Qy 1209 ccaacccagagatgtttgaacctcgtcacttctgtgtgtgtgtgtgtgtgtgtgtgtgt 1268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 ccaagcccgagacaccttcaacaccttcaacaccttcaacaccttcaacaccttcaacacct 443

Qy 1269 agtaactacttactccttctcagcaggaacacacacacacacacacacacacacacacac 1328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 ccccttcacgtacttgccttctcgggcccggccacacacacacacacacacacacacacac 383

Qy 1329 cgcagtagctgttttttctcctgaccttcttctacagaaacttttaacctgaaactctgt 1388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 ctgcttgaggccaagtgcactgtctcagctgtctcagctgtctcagctgtctcagctgtct 323

Qy 1389 attgaccccaagg 1401
    ||| ||| |||
Db 322 cctgagaccagg 310

RESULT 9
US-10-105-299-7201/c
; Sequence 7201, Application US/10105299
; GENERAL INFORMATION:
; APPLICANT: Rosen, et. al
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS950
; CURRENT APPLICATION NUMBER: US/10/105,299
; CURRENT FILING DATE: 2002-03-26
; NUMBER OF SEQ ID NOS: 15197
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7201
; LENGTH: 13327
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-105-299-7201

Query Match      2.8%; Score 45.8; DB 1; Length 1530;
Best Local Similarity 44.6%; Pred. No. 0.012;
Matches 225; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

Qy 882 gacttaacttggagctggagacagacacacacacacacacacacacacacacacacacac 941
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 901 gatatgtctgtgtgaaacagagatttcagcagtcagagtgagtgagtgagtgagtgagtg 960

Qy 942 ctgctgaagcaccacagaggtcacagctaaagtccaggaagagattgaacgtgtcatggc 1001
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 961 ctgctaaagagccagagattttcaaaaagcctacagaagaattggatcgagtaattggg 1020

Qy 1002 agaaacccggagccctgcacagcagcagggggccacacatgcccacacagatgctgtgtgt 1061
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1021 cagaatagatgggtacaagaaaagacattccaaatcttctctacatagaggcaatagtc 1080

Qy 1062 cacaggtccagagatatacatcgacctcatccccaccagcctgccccatgcagtgacctgt 1121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1081 aaagagactatgcactgacaccccggtggcaccacaaatgtgtgtgtgtgtgtgtgtgtgt 1140

Qy 1122 gactttaaatccagaactacactcattcccaaggggcacaaccataattacttccctcaact 1181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1141 gactgtaaggtagcagcaggtctacacgtttaagaaaggaaacacagggtccctgtgtgagc 1200

Qy 1182 tctgtgtacatgacaacaaagaatttcccaaccacagagattgttgacctcgtcaccttt 1241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1201 actatggaaagacccctcacattgtggacgagcctgagcgttcaagccggagaggttcc 1260

Qy 1242 ctggatgaaggtgaaattttaagaaagtaact---acttcagcttcttctcagcagga 1298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1261 cacgaaagttccattgatgtttaaaggacatgattttgagcttttgcctatttggagctgg 1320

Qy 1299 aaacggatttgtgtggagagggcctggtcccgcaatgagctgtttttattcctgaccttc 1358
```

Db 1321 agaaggatgcccgggttaataacttggggcttaagggtgattcaagtagcttagctaat 1380
Qy 1359 atttaccagaactttaacctgaaat 1383
Db 1381 ctatatacattgaatttaactggtcat 1405

RESULT 11

PCT-US02-06912-8
: Sequence 8, Application PC/TUS0206912
: GENERAL INFORMATION:
: APPLICANT: University of Kentucky Research Foundation
: TITLE OF INVENTION: Cytochrome P450s and Uses Thereof
: FILE REFERENCE: 07678/100W03
: CURRENT APPLICATION NUMBER: PCT/US02/06912
: PRIOR FILING DATE: 2002-03-08
: PRIOR APPLICATION NUMBER: US 60/274,241
: PRIOR FILING DATE: 2001-03-09
: PRIOR APPLICATION NUMBER: US 60/275,597
: PRIOR FILING DATE: 2001-03-13
: NUMBER OF SEQ ID NOS: 43
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 1693
: TYPE: DNA
: ORGANISM: Nicotiana tabacum
PCT-US02-06912-8

Query Match 2.6%; Score 45.8; DB 1; Length 1693;
Best Local Similarity 51.7%; Pred. No. 0.012;
Matches 104; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
Qy 87 aaactccctctggccacactctctccagtgattggaataatccacagatagatt 145
Db 194 aactccacactgggtccacagcagtcctcaattttgtgaattgctcgaattggc 253
Qy 147 aaggatgtcagcaaatcttaacaaatctctcaaaatctatggccctgtgttcactctg 206
Db 254 gactgaaccataactcttgccaccatgtcacaaactatgtctctattttactc 313
Qy 207 tatttggcctggaacgcatggtggtctgcatggaatgaagtgggtgaaggaagccctg 265
Db 314 aaacttgggtcaaaacactagctgtggtatcgaaacctgagtagctaaacaaagtctta 373
Qy 267 attgatcttggagaggattt 287
Db 374 cacacgcaagggggtcgagttt 394

RESULT 12

US-09-975-254-18019
: Sequence 18019, Application US/09975254
: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: Heck, Gregory R.
: APPLICANT: La Rosa, Thomas J.
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
: FILE REFERENCE: 38-21(15309)B
: CURRENT APPLICATION NUMBER: US/09/975,254
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US/09/263,191
: PRIOR FILING DATE: 1999-03-05
: NUMBER OF SEQ ID NOS: 31255
: SEQ ID NO 18019
: LENGTH: 254
: TYPE: DNA
: ORGANISM: Glycine max
: OTHER INFORMATION: Clone ID: 700955037H1
US-09-975-254-18019

Query Match 2.6%; Score 44.6; DB 5; Length 254;
Best Local Similarity 50.2%; Pred. No. 0.011;
Matches 110; Conservative 0; Mismatches 109; Indels 0; Gaps 0;
Qy 1035 acatgccctacacagatgctgtggcgacgaggtccagagatacatcgacctcatcccca 1095
Db 6 aaatgccatacctacagctgtgtggaaggagcttctaagaagcaccctccaacact 65
Qy 1096 ccagcctgccccatgcagtgacctgtgacgtttaaatccagaacactacattcccaagg 1155
Db 66 ttgtgctaacaatgctgtgactgagccaccactttggagggtgatgacataccaattg 125
Qy 1156 gcacaaccattataacttccctcacttctgtctacatgacacaaagaatttcccaacc 1215
Db 126 atgcaaatgttaggtgtacacaccagccatgtcgtgaggagaccccaaaaaattgtttaacc 185
Qy 1216 caagatgtttgacccctcacttcttctggtgaggaaggtg 1254
Db 186 ctgagaagtttgacctgagagattcatctctgggggtg 224

RESULT 13

US-10-103-313-626
: Sequence 626, Application US/10103313
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P207C1
: CURRENT APPLICATION NUMBER: US/10/103,313
: CURRENT FILING DATE: 2002-03-12
: NUMBER OF SEQ ID NOS: 653
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 626
: LENGTH: 832
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-103-313-626

Query Match 2.5%; Score 43.6; DB 6; Length 832;
Best Local Similarity 58.6%; Pred. No. 0.033;
Matches 95; Conservative 0; Mismatches 64; Indels 3; Gaps 1;
Qy 823 agaaaagcaaaacacagctctgaattcactattgaaaacttggtaactcactgcagctg 882
Db 627 agtaacagcagggaccggagagccatttccaggagtagacgtcggtaatgacgacgatt 686
Qy 883 acttacttgagctggagacagacacaaagcacaacccctgagatgtgctctctctcc 942
Db 687 tttttttgg---cgtcaccgaaacacagagaccacctgtgctatgggtcctctcattc 743
Qy 943 tgcagaagcaccagaggtcacagctaaagtccagggaagaga 984
Db 744 tgcctaagctacctagaggtgagaggtgtgcaaccggagaga 785

RESULT 14

US-10-103-313-625
: Sequence 625, Application US/10103313
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: P207C1
: CURRENT APPLICATION NUMBER: US/10/103,313
: CURRENT FILING DATE: 2002-03-12
: NUMBER OF SEQ ID NOS: 653
: Prior Application removed - See File Wrapper or Palm
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 625
: LENGTH: 1001
: TYPE: DNA

Search completed: April 19, 2002, 11:06:02
Job time: 10682 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:15:21 ; Search time 277.57 seconds
(without alignment's)
2294.890 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcgaataatttgagcctg.....ttagctcatgtgaacgggg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAI.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAI.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAI.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAI.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAI.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAI.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAI.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAI.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAI.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAI.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAI.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAI.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAI.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAI.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAI.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAI.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAI.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAI.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAI.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAI.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAI.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAI.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed. and is derived by analysis of the total score distribution...

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	622.6	83.8	8437	AAD12242	Human cytochrome P
2	433.6	58.4	1001	AAH51158	Human CYP2C8 relat
3	421.8	56.8	1001	AAH51159	Human CYP2C8 relat
4	415.4	55.9	436	AAH51164	Human CYP2C9 relat
5	369	49.7	734	AAD12247	Human cytochrome P
6	366.2	49.3	557	AAD12245	Human cytochrome P
7	193.2	26.0	570	AAD12249	Human cytochrome P
8	163	21.9	1419	AAH51165	Human cytochrome P
9	163	21.9	1473	AAH51165	Human cytochrome P
10	163	21.9	1473	AAH51165	Human cytochrome P
11	163	21.9	1473	AAH51165	Human derived cyto

12	163	21.9	1591	20	AAH51165	Human cytochrome P
13	163	21.9	1845	20	AAH51165	Mammalian cytochro
14	163	21.9	1852	19	AAH51165	Cytochrome P450 2C
15	163	21.9	1852	19	AAH51165	Human cytochrome P
16	163	21.9	1854	17	AAH51165	Cytochrome P450 2C
17	163	21.9	1854	19	AAH51165	Human cytochrome P
18	163	21.9	1854	19	AAH51165	Partial sequence o
19	161.4	21.7	1818	21	AAH51165	Human liver cytoch
20	153.4	20.6	1892	19	AAH51165	Human cytochrome P
21	150.6	20.3	169	22	AAH51165	Human cytochrome P
22	149.6	20.1	1419	22	AAH51165	Human derived cyto
23	149.6	20.1	1438	22	AAH51165	Human derived cyto
24	149.6	20.1	1447	19	AAH51165	Cytochrome P450 2C
25	149.6	20.1	1473	16	AAH51165	Human auxillary cy
26	149.6	20.1	1473	17	AAH51165	Human cytochrome P
27	149.6	20.1	1669	19	AAH51165	Wild type cytochro
28	149.6	20.1	1746	17	AAH51165	Cytochrome P450 2C
29	149.6	20.1	1746	19	AAH51165	Human cytochrome P
30	149.6	20.1	1748	22	AAH51165	Human cytochrome P
31	149.6	20.1	8475	22	AAH51165	Genetic construct
32	144	19.4	200	16	AAH51165	Human cytochrome P
33	140.2	18.9	1473	16	AAH51165	Human auxillary cy
34	140.2	18.9	1473	17	AAH51165	Human cytochrome P
35	140.2	18.9	2009	17	AAH51165	Cytochrome P450 2C
36	140.2	18.9	2009	19	AAH51165	Human cytochrome P
37	140.2	18.9	2258	17	AAH51165	Cytochrome P450 2C
38	140.2	18.9	2258	19	AAH51165	Human cytochrome P
39	140	18.8	167	22	AAH51165	Human cytochrome P
40	135.8	18.3	1473	16	AAH51165	Human auxillary cy
41	135.8	18.3	1473	17	AAH51165	Human cytochrome P
42	135.8	18.3	1473	17	AAH51165	Human derived cyto
43	135.8	18.3	1922	22	AAH51165	Human liver cell s
44	134.2	18.1	1473	16	AAH51165	Human auxillary cy
45	134.2	18.1	1473	16	AAH51165	Human auxillary cy

ALIGNMENTS

RESULT 1
AAD12242
ID AAD12242 standard; DNA: 8437 BP.
XX
AC AAD12242:
XX
DT 25-SEP-2001 (first entry)
XX
DE Human cytochrome P450 (CYP450) 2C19 gene.
XX
KW Human: gene structure: phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH misc_signal 700..702
FH primer_bind /*tag= a
FH primer_bind /*note= "Translation start site"
FH primer_bind 818..833
FH primer_bind /*tag= b
FH primer_bind /*bound_moiety= "Primer"
FH primer_bind 1535..1557
FH primer_bind /*tag= c
FH primer_bind /*bound_moiety= "Primer"
FH primer_bind 1578..1597
FH primer_bind /*tag= d
FH primer_bind /*bound_moiety= "Primer"
FH primer_bind 1861..1882
FH primer_bind /*tag= e
FH primer_bind /*bound_moiety= "Primer"
FH primer_bind 1879..1898
FH primer_bind /*tag= f

FT	primer_bind	/bound_moiety=	"Primer"
FT		3147..3168	
FT		/*tag= g	
FT	primer_bind	/bound_moiety=	"Primer"
FT		3258..3279	
FT		/*tag= h	
FT	primer_bind	/bound_moiety=	"Primer"
FT		3460..3479	
FT		/*tag= i	
FT	primer_bind	/bound_moiety=	"Primer"
FT		4876..4890	
FT		/*tag= j	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6049..6064	
FT		/*tag= k	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6091..6105	
FT		/*tag= l	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6093..6112	
FT		/*tag= m	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6858..6872	
FT		/*tag= n	
FT	primer_bind	/bound_moiety=	"Primer"
FT		6874..6893	
FT		/*tag= o	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7808..7827	
FT		/*tag= p	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7810..7831	
FT		/*tag= q	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7937..7956	
FT		/*tag= r	
FT	primer_bind	/bound_moiety=	"Primer"
FT		7940..7961	
FT		/*tag= s	
FT	primer_bind	/bound_moiety=	"Primer"
FT			

WO200153529-A2.

26-JUL-2001.

17-JAN-2001: 2001WO-US01461.

20-JAN-2000- 2000US-0488127

(GENO-) GENOME THERAPEUTICS CORP.

Thomann H. Fitzgerald ms.

WPT: 2001-465380/50

Determining structure of genes whose sequence is not known from cDNA, by sequencing the gene or gene across exon-intron boundaries using evenly spaced primers comprising nucleic acids that hybridize to the cDNA of gene -

Claim 23: Fig 5: 81pp; English.

The present invention relates to a method for determining gene structure when the genomic sequence is unknown. The method involves sequencing the gene across exon-intron boundaries using evenly spaced primers or tiled primers. The tiled primers comprises nucleic acids that hybridise to the known cDNA sequence of the gene at about 170-300 base intervals and the gene comprises the template. Gene structure can be determined without the need to sequence the entire gene. The method provides information necessary to determine gene structure and phenotypic expression without the need to sequence entire chromosomal copy of the gene or fragment. The methods are useful in germline sequence variation analysis. The method is also useful for determining the boundaries

CC between regions of nucleic acids that were separated by intervening
CC sequence, and also for determining boundaries present in genes containing
CC group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
CC in the presence of guanosine cofactor. The present sequence is human
CC cytochrome P450 (CYP450) 2C19 gene related to the invention.
XX
XX
SQ Sequence 8437 BP: 2392 A: 1501 C: 1541 G: 2654 T: 349 other:

Query Match	83.8%	Score 622.6;	DB 22;	Length 8437;
Best Local Similarity	93.3%	Pred. NO. 1.3e-183;		
Matches 694;	Conservative	0;	Mismatches 45;	Indels 5;
				Gaps 4;

Oy	1	tcaaaaattttgaagccgttgtgcttgaaataaaagcatacaaaacacaatgaaaaatacca	50
Db			
Db	1391	tcaaaaaattttg-agcctgtgtgactgaataaaagcatacaaaacacaatgaaaaatacca	1449
Oy	61	tgttaaatcaaggcttagcaaaatggacaaaaatagtaaacttctgtctgttatctctgtct	120
Db			
Db	1450	atctaagtcaagcttagtaaatggacaaaacagtgaacttcattctgtctttaactgatct	1509
Oy	121	accttcttagcttcaaagggtctatggcccctgtgttcaactctgtattttggcctgaaacc	180
Db			
Db	1510	ccctttcnag-tctcaaaaatataggcccctgtgttcaactctgtattttggcctggaacg	1568
Oy	181	catagtgtctgcattgatataagaagcagtgaaagaaacctgattgatcttggaggaga	240
Db			
Db	1569	cattggtgtgtcgatggaatagaagtggtagaaggagccctgattgatcttggaggaga	1628
Oy	241	gttttctggaagagacattttcccactggctgaaaagactaacaaagagaatttggtaggtg	300
Db			
Db	1629	gttttctggaagagccattttcccactggctgaaaagactaacaaagagaatttggtaggtg	1688
Oy	301	tgcattgtccctgtttcacgacatctgtcttgggagatgggagagatggaasaacagagacttac	360
Db			
Db	1689	tqcaagtccctgtttcacgacatctgtcttgggagatgggagagatggaasaac--agactagc	1745
Oy	361	agagctctctggcgagagcttggccccatccacatggctgccagtgctcagettctcttt	420
Db			
Db	1747	agagcttctggcgagagcttggccccatccacatggctgccagtgctcagettctcttt	1806
Oy	421	cttgcctgggattccctctcagtttctcttctctctctgttagaaattgttttcagcaa	480
Db			
Db	1807	cttgcctgggattccctctcagtttctcttctctctctgttagaaatcgtttttcagcaa	1866
Oy	481	tggaaagaaatggaaagagatccgcgctttctccctcatgacgtgcggaattttgggat	540
Db			
Db	1857	tggaaagagatggaagagatccgcgctttctccctcatgacgtgcggaattttgggat	1926
Oy	541	gggggaagaggagcattgaggaccgttgttcaagaggaaagcccgcctgcttggaggagtt	600
Db			
Db	1927	gggggaagaggagcattgaggaccgttgttcaagaggaaagcccgcctgcttggaggagtt	1986
Oy	601	gagaaaaaccaagggtgggtgacctactccatatcactgaccttacttggagctactct	660
Db			
Db	1987	gagaaaaaccaagggtgggtgaaacatactctatacactgacctttc-ggagctgctctcc	2045
Oy	661	tctctactgacattcttggaaacatttcagggggtggccatacttcttacttatgagt-ctg	719
Db			
Db	2047	tctctactgacattcttggaaacatttcagggggtggccagatcttttatttggagctctg	2106
Oy	720	gttgttaactcatgtgaagcggggg	743
Db			
Db	2107	gttgttaactcatgtgaagcaggg	2130

RESULT 2

AAH51158
ID AAH

XXXXXXXXXXXXXXXXXXXX

XX

IN 22-AUG-2007 (JSTJ) (JSTJ)


```
XX PF 24-MAR-2000: 2000WO-IB00403.
XX PR 25-MAR-1999: 99US-0126269.
XX PR 30-APR-1999: 99US-0131961.
XX PA (GEST ) GENSET.
XX PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A:
XX DR WPI: 2000-638353/61.
XX PT Polynucleotides comprising sequences from malate decarboxylase
XX PI enzyme-related biallelic markers used for genotyping -
XX PS Claim 13: Page 280; 673pp; English.
XX CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
XX CC biallelic markers. The sequences are related to various human genes
XX CC including microsomal glutathione S-transferase II (MGSTII), malate
XX CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
XX CC reductase/synthase (GSHR/GSHS), flavin-containing monooxygenase (FMO),
XX CC gamma-glutamyltransferase 5 (GGT5), dipeptidase (DP), glucose 5-phosphate
XX CC dehydrogenase (G6PDH), phosphogluconate dehydrogenase (PGDH), and uridine
XX CC diphosphate glucuronosyl transferases (UGT2). Each of these sequences
XX CC contains a biallelic marker/polymorphism, which is represented in the
XX CC sequence as a degenerate/undefined base. The genes to which the biallelic
XX CC marker containing sequences are related are involved in drug metabolism.
XX CC Sequences AAH51594 - AAH51598 represent the genomic sequence of the
XX CC MGSTII gene and four alternative MGSTII cDNA sequences. AAH52905-AAH52906
XX CC are MGSTII gene products. PCR primers AAH51599 and AAH51500 are used in
XX CC an example for the amplification of human genomic DNA fragments. The
XX CC invention includes a method of genotyping comprising determining the
XX CC identity of a nucleotide at a DME- or MGSTII-related biallelic marker in
XX CC a biological sample. The method is used to determine the frequency in
XX CC a population of an allele of a DME- or MGST-II related biallelic marker and
XX CC to select an individual for inclusion in a clinical trial of a drug
XX CC treatment. The method is also used to detect association between allele
XX CC and phenotype, and to detect association between haplotype and phenotype.
XX CC The polynucleotides are used, in hybridization assays, sequencing assays
XX CC or allele specific amplification assays. The method can be used to
XX CC determine whether an individual suffers or is at risk of developing
XX CC asthma or is at risk of developing hepatotoxicity on treatment with
XX CC zileuton.
XX SO Sequence 1001 BP: 265 A; 188 C; 244 G; 302 T; 2 other;
```

```
Query Match 56.8%; Score 421.8; DB 21; Length 1001;
Best Local Similarity 78.0%; Pred. No. 2.1e-121;
Matches 544; Conservative 1; Mismatches 148; Indels 4; Gaps 3;
```

```
Qy 49 atgaataatcatcgtctaaatcaggcttagcaaatgagcaaaaatagtaacttccttgcctg 108
Db 1 atgaataatcagctgaatcatcatagatttgtagcaaatagcgaacttatttgcctg 60
Qy 109 ttatctgtctacttctcctagctctcaaaaggctctatggccctgtgttcaactcttatt 168
Db 61 ctattgcatcttctccctcagttctcaaaagtctatggtccctgtgttccacgttatt 120
Qy 169 tggcctgaaccataatggtgtgtgatgatgatgaagcagtagtaagaagccctattga 228
Db 121 tggcatgaatcccatagtggtttctgatgatgatgagcagtagtaagaagccctattga 180
Qy 229 tcttgaggagagatttctgaaagagcatttcccaactggtgaaagagcagtaaragag 288
Db 181 taatggagagagatttctgaaagagcagtaaragagcagtaaragagcagtaaragag 240
Qy 289 atttgtaggtgtagtgcctgttctcagcattctgtcttgaggatgggagagatggaaa 348
Db 241 acttggtaggtgacacatttctgtgtcagcttggtaactggggtgaggggagatggaaa 300
Qy 349 acagagacttacagagctctctcggcgagagcttggcccattccacatggtgcccctagt 407
```

```
Db 301 acagagcctaaaagcttctc-agcagagcttagcctatctgcatgctgctccagtggtt 359
Qy 408 -cagcttctcttcttctgctgggagctcctcctctctagttctgttcttctctcttagga 466
Db 350 gcagacacttcttctgctggctggaattctc-ccagtttctgcccctttttttattagga 418
Qy 457 attgttttcagcaatgaaagaaatggaagagatccggcgtttcttctcctcatgacgtg 526
Db 419 atcatttccagcaatgaaagagatggaagagatccggcgtttcttctcctcacaccttg 478
Qy 527 cggaaatttggatgggaagagcattgagacccgttgcagaccgtgttcaagaggaagcccgctgc 586
Db 479 cggaaatttggatgggaagagcattgagacccgttgcagaccgtgttcaagaggaagcctcgc 538
Qy 587 ctgtgagagaggttgagaaaaaaccaagaggtgggtgagaccctactcattacactaccta 646
Db 539 ctgtgagagaggttgagaaaaaaccaagaggtgggtgagactctactctgcctcattgacctta 598
Qy 647 ctggactactcttcttactgacattcttggaaacatttcagggggtggccattcttt 706
Db 599 acagtactctgtcttactagtagcagctctctggaaaacatttcagggggtggccaggtcttc 658
Qy 707 cattatgactctggtttagctcatgtgaaagcgggg 743
Db 659 attgcacatcctggttgcagccctcaggtggtgag 695
RESULT 4
AAH51164
ID AAH51164 standard: DNA; 436 BP.
AC AAH51164;
DT 29-AUG-2001 (first entry)
DE Human CYP2C9 related DNA containing a biallelic polymorphism SEQ ID 55.
KW Human: biallelic marker; single nucleotide polymorphism; SNP; MGSTII;
KW microsomal glutathione S-transferase II; malate decarboxylase enzyme;
KW DME1; ME1; cytochrome P450; glutathione reductase; GSHR; GSHS; GGT5;
KW flavin-containing monooxygenase; FMO; gamma-glutamyltransferase 5;
KW dipeptidase; DP; glucose 6-phosphate dehydrogenase; G6PDH; haplotype;
KW phosphogluconate dehydrogenase; PGDH; drug metabolism; phenotype;
KW uridine diphosphate glucuronosyl transferase; UGT2; asthma; hepatotoxicity;
KW zileuton; ds.
OS Homo sapiens.
XX
XX WO2000058508-A2.
XX
XX 05-OCT-2000.
XX
XX 24-MAR-2000: 2000WO-IB00403.
XX
XX 25-MAR-1999: 99US-0126269.
XX PR 30-APR-1999: 99US-0131961.
XX (GEST ) GENSET.
XX
XX Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A:
XX WPI: 2000-638353/61.
XX Polynucleotides comprising sequences from malate decarboxylase
XX PI enzyme-related biallelic markers used for genotyping -
XX PS Claim 13: Page 284; 673pp; English.
XX CC Sequences AAH51110-AAH51593 represent human DNA fragments which contain
XX CC biallelic markers. The sequences are related to various human genes
XX CC including microsomal glutathione S-transferase II (MGSTII), malate
XX CC decarboxylase enzyme (DME1/ME1), cytochrome P450, glutathione
```

XX Human: gene structure; phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX
OS Homo sapiens.
XX
PN W0200153529-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US01461.
XX
PR 20-JAN-2000; 2000US-0488127.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Thomann H, Fitzgerald MS;
XX WPI: 2001-465380/50.
XX
XX Determining structure of genes whose sequence is not known from cDNA.
PT by sequencing the gene or gene across exon-intron boundaries using
PT evenly spaced primers comprising nucleic acids that hybridize to the
PT cDNA of gene
XX
XX Example 2; Fig 4; 8lpp; English.
XX
XX The present invention relates to a method for determining gene
CC structure when the genomic sequence is unknown. The method involves
CC sequencing the gene across exon-intron boundaries using evenly spaced
CC primers or tilted primers. The tiled primers comprises nucleic acids that
CC hybridize to the known cDNA sequence of the gene at about 100-300 base
CC intervals and the gene comprises the template. Gene structure can be
CC determined without the need to sequence the entire gene. The method
CC provides information necessary to determine gene structure and phenotypic
CC expression without the need to sequence entire chromosomal copy of the
CC gene or fragment. The methods are useful in germline sequence variation
CC analysis. The method is also useful for determining the boundaries
CC between regions of nucleic acids that were separated by intervening
CC sequence, and also for determining boundaries present in genes containing
CC group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
CC in the presence of guanosine cofactor. The present sequence is a PCR
CC product obtained from human cytochrome P450 (CYP450) 2C19 gene
CC and 349L primer related to the invention.
XX
XX Sequence 734 BP; 236 A; 170 C; 126 G; 199 T; 3 other;
SQ

Query Match 49.7%; Score 369; DB 22; Length 734;
Best Local Similarity 90.7%; Pred. No. 5,3e-105;
Matches 427; Conservative 0; Mismatches 40; Indels 4; Gaps 3

Qy 1 tcagaaatatttgaagccgtgtgtggtcgaataaaagcgcatacaaatcaaatgaaataatca 60
Db 458 TCAGAAATATTTGAGCGCTGTGTGACACTTGAATATAAGCATAACAATACAAATGATA 409
Qy 61 tgcataatcagccttagcaaatagcacaataagtaacttcgttgcgtatctctgct 120
Db 408 ATCTAGTCAGGCTTATAGTAATGG-CAAAACAGTGACTTCATTCGCTGTTAACTGTATCT 350
Qy 121 accttctagctctcaaggtctatggccctgtgttccactctgtatcttggcctgaaacc 180
Db 349 CCTTTCTAGCTCTCAAAAATCATGGCCCTGTGTTCACCTCTGTATTTTGGCCTGGAACG 290
Qy 181 catagtgtctcatgatataagcaatgaagaagccctgattgatcttgaggaga 240
Db 289 CATGGTGGTCTCATGGATATGAAGTGTGAAGGANGCCCTGATTTGATCTTGGAGAGA 230
Qy 241 gtttctggaagaggcatcttcccactgctgaaagagctaacagagatttggttagtg 300
Db 229 GTTTCTGGAAGGCCATTTCCTACTGCTGAACAGCTAACACAGGATTTGTAGGTG 170

[illegible]

```

QY 301 tgcattgctgttttcagcatctgtcttggggatgggagagatgggaaacagagctttac 360
Db 169 TGAAGTGGCTGTTTCAGCATCTGCTTGGGATGGGAGGATGGAAAC--AGACTAGC 112
QY 361 agagctctcgggcagagcttggcccatccacatgctgcccagctgcagcttctcttt 420
Db 111 AGAGCTTCTCGGCAGAGCTTGGCCCATCCACATGCTGCCAGTGTGACGTTCTCTTT 52
QY 421 ctgctcgggagatccctcctcagtttgcttctctt-cctgttaggaattg 47
Db 51 CTTGCTGGGATCTCCCTCCTAGTTTCGTTCTCTTCTGCTGTAGGAATCG 1

RESULT 6
AADI2245
ID AADI2245 standard; DNA: 557 BP.
XX
AC AADI2245;
XX
XX 25-SEP-2001 (first entry)
XX
DE Human cytochrome P450 (CYP450) 2C19 gene PCR product #2.
XX
XX Human; gene structure: phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX
XX Homo sapiens.
XX
XX WO200153529-A2.
XX
XX 26-JUL-2001.
XX
XX 17-JAN-2001; 2001WO-US01461.
XX
XX 20-JAN-2000; 2000US-0488127.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Thomann H, Fitzgerald MS;
XX
XX WPI: 2001-465380/50.
XX
XX Determining structure of genes whose sequence is not known from cDNA,
XX by sequencing the gene or gene across exon-intron boundaries using
XX evenly spaced primers comprising nucleic acids that hybridize to the
XX cDNA of gene.
XX
XX Example 2; Fig 4; 81pp; English.
XX
XX The present invention relates to a method for determining gene
XX structure when the genomic sequence is unknown. The method involves
XX sequencing the gene across exon-intron boundaries using evenly spaced
XX primers or tiled primers. The tiled primers comprises nucleic acids that
XX hybridize to the known cDNA sequence of the gene at about 100-300 base
XX intervals and the gene comprises the template. Gene structure can be
XX determined without the need to sequence the entire gene. The method
XX provides information necessary to determine gene structure and phenotypic
XX expression without the need to sequence entire chromosomal copy of the
XX gene or fragment. The methods are useful in germline sequence variation
XX analysis. The method is also useful for determining the boundaries
XX between regions of nucleic acids that were separated by intervening
XX sequence, and also for determining boundaries present in genes containing
XX group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
XX in the presence of guanosine cofactor. The present sequence is a PCR
XX product obtained from human cytochrome P450 (CYP450) 2C19 gene
XX and 235U primer related to the invention.
XX
XX Sequence 557 BP; 119 A; 107 C; 137 G; 167 T; 27 other;

```

Query Match 49.9%; Score 366.2; DB 22; Length 557;
 Best Local Similarity 89.4%; Pred. No. 3.4e-104;

```

Matches 429; Conservative 0; Mismatches 45; Indels 6; Gaps 4:
QY 219 cccctgattgatttgcagagagatgtttcttggaaagagcattttccacatggctgaagaag 278
Db 2 cccctgattgatttgcagagagatgtttcttggaaagagcattttccacatggctgaagaag 61
QY 279 ctacacagagatttgcagagagatgttcctgttccagcatctgtcttgggagtgagg 338
Db 62 ctacacagagatttgcagagagatgttcctgttccagcatctgtcttgggagtgagg 121
QY 339 agatggaacacagagacttacagagatcctcgggcagagcttggcccatccacatggct 398
Db 122 aggatggaaac--agactagcagagcttctcgggcagagcttggcccatccacatggct 179
QY 399 gccagtgcaactcctcttcttctccttgggagatcctcctcctagtttctgttcttccc 458
Db 180 gccagtgcaactcctcttcttctccttgggagatcctcctcctagtttctgttcttccc 239
QY 459 ttttaggaattgttttcagcaatggaagaaatggaagagagatccggcttttctcctca 518
Db 240 ttttaggaatcgttttcagcaatggaagaaatggaagagagatccggcttttctcctca 299
QY 519 tgacgtcgcgaatttgggatgggagagagagagagagagagagagagagagagagagag 578
Db 300 tgacgtcgcgaatttgggatgggagagagagagagagagagagagagagagagagagag 359
QY 579 cccgtccttgggagagagagagagagagagagagagagagagagagagagagagagag 636
Db 360 ccc-cttctctgtgganganttgagaaacaaatgggtgggnaacatactntttatc 418
QY 637 actgaccttac-tggactactatcttctctactgacattcttggaaacatttcaagggatg 695
Db 419 actgaccttcttgnactntntnctnttactgacattttnnggaaacatttnangngg 478

RESULT 7
AADI2249
ID AADI2249 standard; DNA: 570 BP.
XX
AC AADI2249;
XX
XX 25-SEP-2001 (first entry)
XX
DE Human cytochrome P450 (CYP450) 2C19 gene PCR product #4.
XX
XX Human; gene structure: phenotypic expression; guanosine cofactor;
KW germline variation analysis; exon-intron boundary; Tetrahymena rRNA;
KW cytochrome P450 2C19; CYP450 2C19; ds.
XX
XX Homo sapiens.
XX
XX WO200153529-A2.
XX
XX 26-JUL-2001.
XX
XX 17-JAN-2001; 2001WO-US01461.
XX
XX 20-JAN-2000; 2000US-0488127.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Thomann H, Fitzgerald MS;
XX
XX WPI: 2001-465380/50.
XX
XX Determining structure of genes whose sequence is not known from cDNA,
XX by sequencing the gene or gene across exon-intron boundaries using
XX evenly spaced primers comprising nucleic acids that hybridize to the
XX cDNA of gene.
XX
XX Example 2; Fig 4; 81pp; English.
XX
XX The present invention relates to a method for determining gene
XX structure when the genomic sequence is unknown. The method involves
XX sequencing the gene across exon-intron boundaries using evenly spaced
XX primers or tiled primers. The tiled primers comprises nucleic acids that
XX hybridize to the known cDNA sequence of the gene at about 100-300 base
XX intervals and the gene comprises the template. Gene structure can be
XX determined without the need to sequence the entire gene. The method
XX provides information necessary to determine gene structure and phenotypic
XX expression without the need to sequence entire chromosomal copy of the
XX gene or fragment. The methods are useful in germline sequence variation
XX analysis. The method is also useful for determining the boundaries
XX between regions of nucleic acids that were separated by intervening
XX sequence, and also for determining boundaries present in genes containing
XX group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
XX in the presence of guanosine cofactor. The present sequence is a PCR
XX product obtained from human cytochrome P450 (CYP450) 2C19 gene
XX and 235U primer related to the invention.
XX
XX Sequence 557 BP; 119 A; 107 C; 137 G; 167 T; 27 other;

```

Query Match 49.9%; Score 366.2; DB 22; Length 557;
 Best Local Similarity 89.4%; Pred. No. 3.4e-104;

CC structure when the genomic sequence is unknown. The method involves
CC sequencing the gene across exon-intron boundaries using evenly spaced
CC primers or tiled primers. The tiled primers comprises nucleic acids that
CC hybridise to the known cDNA sequence of the gene at about 100-300 base
CC intervals and the gene comprises the template. Gene structure can be
CC determined without the need to sequence the entire gene. The method
CC provides information necessary to determine gene structure and phenotypic
CC expression without the need to sequence entire chromosomal copy of the
CC gene or fragment. The methods are useful in germline sequence variation
CC analysis. The method is also useful for determining the boundaries
CC between regions of nucleic acids that were separated by intervening
CC sequence, and also for determining boundaries present in genes containing
CC group 1 type introns such as Tetrahymena rRNA, where self-splicing occurs
CC in the presence of guanosine cofactor. The present sequence is a PCR
CC product obtained from human cytochrome P450 (CYP450) 2C19 gene
CC
XX
SQ Sequence 570 BP; 146 A; 101 C; 142 G; 178 T; 3 other;

Query Match 26.0%; Score 193.2; DB 22; Length 570;
Best Local Similarity 93.8%; Pred. No. 3.4e-50;
Matches 212; Conservative 0; Mismatches 13; Indels 1; Gaps 1;
QY 519 tgacgtcgcgaatttggatggggaagagagcattgagaccgtgttcaagagggaag 578
DB 4 tgacgtcgcgaatttggatggggaagagagcattgagaccgtgttcaagagggaag 63
QY 579 ccgcgtccttggaggaggttgagaaaaaaccaagggtgggtgacctactctcatcac 638
DB 64 ccgcgtccttggaggaggttgagaaaaaaccaagggtgggtgacctactctcatcac 123
QY 639 tgacctactgactactatctctactgacattctcttggaacatttcagggggtgcc 698
DB 124 tgaccttctgactgctctctctactgacattctcttggaacatttcagggggtgcc 187
QY 699 atattcttcatatgagt-ctggtgttagctcatgtgaagcgggg 743
DB 184 agatctttattggagtcctgtgttagctcatgtgaagcagg 223

RESULT 8
AA087716
ID AAX22716 standard; DNA: 1419 BP.
XX
AC AAX22716;
XX
DT 27-MAY-1999 (first entry)
XX
DE Human cytochrome P450 2C10 variant DNA.
XX
KW Cytochrome P450; detergent; enzyme inhibitor; non-ionic; ionic;
KW purification; drug oxidation; steroid; carcinogen; pesticide; human; ss.
XX
OS Homo sapiens.
XX
PN US5886157-A.
XX
PD 23-MAR-1999.
XX
PF 10-FEB-1994; 94US-0194981.
XX
PR 10-FEB-1994; 94US-0194981.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Gillam EMJ, Guengerich FP, Guo Z, Sandhu P;
XX WPI: 1999-228609/19.
XX
DR Purifying recombinant cytochrome P450 - b; utilising novel
PT combinations of detergents and enzyme inhibitors
XX

PS Disclosure: Column 55-56; 9lpp; English.
XX
CC This invention describes a recombinant cytochrome P450 protein which
CC is purified from a host cell culture using a combination of detergents
CC and enzyme inhibitors. The method comprises (a) fractionating the host
CC cells to prepare their membranes, (b) adding a non-ionic detergent to
CC the membranes in a concentration of 0.8% to 2% (w/v) in a
CC detergent:protein ratio of between 4:1 to 10:1, (c) adding an ionic
CC detergent to the membranes in a concentration of 0.4% to 0.8% (w/v) in a
CC detergent:protein ratio of between 2:1 to 4:1 (d) centrifuging the
CC membrane-detergent mixture to remove insoluble materials and (e)
CC purifying the protein through a diethylaminoethyl-beaded column, then
CC through a carboxymethyl-beaded column, and finally through a
CC hydroxylapatite column. The method is used to purify cytochrome P450
CC proteins which are responsible for catalysing the oxidation of drugs,
CC steroids, carcinogens and pesticides. The method simplifies the
CC purification of P450 proteins, by the use of improved expression vectors
CC and novel detergent combinations.
XX
SQ Sequence 1419 BP; 409 A; 332 C; 307 G; 371 T; 0 other;

Query Match 21.9%; Score 163; DB 20; Length 1419;
Best Local Similarity 97.1%; Pred. No. 1.5e-40;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 126 cctaactctcaaggctatggccctgtgttcactctgtatttggcctgaacccataag 185
DB 110 ccaactctcaaggctatggccctgtgttcactctgtatttggcctgaacccataag 169
QY 186 tggctgctcatggatgaagcagtgaaagagccctgattgatttgaggagaggttt 245
DB 170 tggctgctcatggatgaagcagtgaaagagccctgattgatttgaggagaggttt 229
QY 246 ctggaagaggcatttcccactggctgaaagagctaacagagaggtttgta 296
DB 230 ctggaagaggcatttcccactggctgaaagagctaacagagaggtttgta 280

RESULT 9
AA087715
ID AA087715 standard; cDNA: 1473 BP.
XX
AC AA087715;
XX
DT 10-NOV-1995 (first entry)
XX
DE Human cytochrome P450 molecular species 2C9 cDNA.
XX
KW Human cytochrome P450; amplification; PCR; primer; expression vector;
KW Yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KW carcinogen; mutagen; liver metabolism; ds.
XX
OS Homo sapiens.
XX
PN EP644267-A.
XX
PD 22-MAR-1995.
XX
PF 20-JUL-1994; 94EP-0111298.
XX
PR 21-JUL-1993; 93JP-0180246.
XX
PR 20-JUL-1993; 93JP-0201120.
XX
PR 30-JUL-1993; 93JP-0208279.
XX
PA (HAYA/) HAYASHI K.
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX
DR WPI: 1995-116991/16.
DR P-PSDB: AAR72361.


```

PA (ROMK/) ROMK-SPARKS M.
XX Goldstein JA, Romkes-sparks M;
XX WPI: 1998-436528/37.
XX
XX Screening for drugs metabolised by cytochrome P450 - for identifying
XX mutagenic, carcinogenic, or cytotoxic compounds
XX
XX Example 2: Column 53-56: 63pp; English.
XX
XX This sequence encodes a human cytochrome P450 2C9 polypeptide isolated
XX from clone 65. This polypeptide is used in a method to screen for a drug
XX that is metabolised by a cytochrome P450 having S-mephenytoin
XX 4'-hydroxylase activity. The protein can also be used to identify a
XX mutagenic, carcinogenic or cytotoxic compound. Cytochrome P450 are a
XX large family of haemoprotein enzymes capable of metabolising xenobiotics
XX such as drugs carcinogens and environmental pollutants as well as
XX endobiotics such as steroids, fatty acids and prostaglandins.
XX
SQ Sequence 1852 BP: 514 A: 424 C: 380 G: 534 T: 0 other:

Query Match      21.9%; Score 163; DB 19; Length 1852;
Best Local Similarity 97.1%; Pred. No. 1.7e-40;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0:

Oy 126 cctagctctcaaaaggctctatggccctgtgttcactctgtattttggccctgaaccaccatag 185
Db 174 ccaatctctcaaaaggctctatggccctgtgttcactctgtattttggccctgaaccaccatag 233

Oy 186 tgggtctgcctgatatgaagcagtgaaagagccctgattgacatcttgaggagagatttt 245
Db 234 tgggtctgcctgatatgaagcagtgaaagagccctgattgacatcttgaggagagatttt 293

Oy 246 ctggaagaggcattttccactggctgaaagagctaacagagagatttggtta 296
Db 294 ctggaagaggcattttccactggctgaaagagctaacagagagatttggtta 344

```

Search completed: April 19, 2002, 08:15:28
Job time: 3228 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:07:44 ; Search time 2723.34 Seconds
(without alignments)
2931.736 Million cell updates/sec

Title: US-09-763-292-2

Perfect score: 743

Sequence: 1 tcgaataatttgaagctg,.....ttagctcatgtgaagcggg 743

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gb_est1:*
11: gb_est2:*
12: gb_hic:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rtd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	324.2	43.6	675	10	AV683520
2	322.8	43.4	628	10	AV693304
3	322.8	43.4	631	10	AV695514
4	322.8	43.4	692	10	AV697749
5	321.8	43.3	651	10	AV697837
6	321.8	43.3	675	10	AV696436
7	321.8	43.3	677	10	AV697538
8	321.8	43.3	677	10	AV698646
9	317	42.7	580	10	AV696639
10	314.8	42.4	622	10	AV697579
11	289.6	39.0	747	10	AV646984
12	280.2	37.7	646	10	AV696485

13	270.4	36.4	721	11	BG617892
14	267.4	36.0	530	10	AV695838
15	253.2	34.1	395	10	AV660643
16	249.8	33.6	608	11	BG563276
17	245.6	33.1	716	11	BG569317
18	225.8	30.4	425	11	T53876
19	221.4	29.8	724	10	AV697373
20	209.8	28.2	667	10	AV697578
21	171	23.0	401	10	AV694577
22	163	21.9	344	10	AV656121
23	163	21.9	346	10	AV661103
24	163	21.9	360	10	AV659393
25	163	21.9	383	10	AV649200
26	163	21.9	385	10	AV649420
27	163	21.9	388	10	AV655432
28	163	21.9	407	10	AV645669
29	163	21.9	445	10	AV649415
30	163	21.9	612	10	AV661158
31	163	21.9	613	10	AV655841
32	163	21.9	615	10	AV685094
33	163	21.9	620	10	AV647944
34	163	21.9	638	10	AV698137
35	163	21.9	667	10	AV653206
36	162	21.8	361	10	AV661559
37	152	21.8	362	10	AV661295
38	152	21.8	364	10	AV661281
39	155.2	20.9	371	10	AV691273
40	155.2	20.9	376	10	AV691272
41	155.2	20.9	390	10	AV688076
42	155.2	20.9	396	10	AV690361
43	155.2	20.9	409	10	AV687011
44	152.4	20.5	363	10	AV655255
45	152.4	20.5	421	10	AV650504

ALIGNMENTS

RESULT 1
AV683520
LOCUS AV683520 675 bp mRNA EST 25-SEP-2000
DEFINITION AV683520 GK Homo sapiens cDNA clone GKCBJD11 5', mRNA sequence.
ACCESSION AV683520
VERSION AV683520.1 GI:10285383
KEYWORDS EST.
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 675)
AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Ou,J., Song,H., Cheng,Z., Zeng,L., Yu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GK- clones
JOURNAL Unpublished (2000)
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source Location/Qualifiers
1..675
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBJD11"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"

```
/note="Vector: pBluescript sk(-): Site_1: EcoRI: Site_2:
XhoI"
BASE COUNT 178 a 138 c 165 g 194 t
ORIGIN

Query Match 43.6%; Score 324.2; DB 10; Length 675;
Best Local Similarity 78.2%; Pred. No. 2e-78;
Matches 415; Conservative 0; Mismatches 113; Indels 3; Gaps 2;

QY 86 caaaatagtaactcgttggctgtatctctgtctactcttctctagctctcaaggcttat 145
   || || || || || || || || || || || || || || || || || || || || ||
Db 5 CACCAATGTAAGTGCCTTAAGTCTCTCCAGCCNAITGCARAGGTCTCAAAAGTCTAT 54

QY 146 ggcctctgtctactctgtatttttggcctgaaaccccatagtgctgcatgcatgaa 205
   || || || || || || || || || || || || || || || || || || || || ||
Db 65 GGTCTGTGTTCACCGGTATTTGGCATGAATCCCATAGTGGTGTTCATGATAGAG 124

QY 206 qcagtgaagaagccctgattgattcttgagagagattttctgaagaagcattttccca 265
   || || || || || || || || || || || || || || || || || || || || ||
Db 125 GCAGTGAAGGAAGCCCTGATGATAATGGAGAGAGATTTTCTGGAAGAGCAATCCCA 184

QY 266 ctggctgaagagagtaacagagatttgtagtgatgctgctgctgtttcagcatctgt 325
   || || || || || || || || || || || || || || || || || || || || ||
Db 185 ATATCTCAAGAATTAATAAGGACTTGGTAGTGACATATTCTGTGTCAGCTTTGGT 244

QY 326 cttgggagtgaggagtgaaacagagacttacagagctctcctgggcagagcctggcc 385
   || || || || || || || || || || || || || || || || || || || || ||
Db 245 AACTGGGGTGAGGGGATGAAACACAGAGCCCTAAAAGCTTCTC-AGCAGAGCTTAGCC 303

QY 386 catcacatggctgcccagtg--cagctctctcttcttgcctggagctctcctcctag 443
   || || || || || || || || || || || || || || || || || || || || ||
Db 304 TAATGATGGCTGCCAGTGTTCAGCAGCTTCTTCCTTGGGTGTGAAATCTCCAGT 363

QY 444 ttctgttctctctctgttaggaattgttttcagcaatgaaagaaatggaagagatcc 503
   || || || || || || || || || || || || || || || || || || || || ||
Db 364 TCTGCCCTTTTATAGGAATCATTTTCAGCAATGGAAGAGATGGAAGGAGATCC 423

QY 504 ggcgttctccctcatagctcagaaatttggatggaggaagaaagacattgaagacc 563
   || || || || || || || || || || || || || || || || || || || || ||
Db 424 GGCCTTCTCCCTCACAACTTCGGAAATTTGGGATGGGAAGAGAGAGCATTCAGGACC 483

QY 564 gtgtcaaggaagagccctgctgtgagagagtggtgaaagaaaccaaagg 614
   || || || || || || || || || || || || || || || || || || || || ||
Db 484 GTGTCAAGGAAGAGCTCACTGCTTGTGGAGGAGTTGAGAAAACCAAGG 534

RESULT 2
AV693304 628 bp mRNA EST 25-SEP-2000
LOCUS AV693304 GK Homo sapiens cDNA clone GKCBIB04 5', mRNA sequence.
DEFINITION AV693304
ACCESSION AV693304
VERSION AV693304.1 GI:10295167
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wu, T., Qian, B., Huang, O., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
, Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
, Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
```

```
Location/Qualifiers
1. 628
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBIB04"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-): Site_1: EcoRI: Site_2:
XhoI"
BASE COUNT 164 a 127 c 160 g 176 t 1 others
ORIGIN

Query Match 43.4%; Score 322.8; DB 10; Length 628;
Best Local Similarity 77.6%; Pred. No. 4.7e-78;
Matches 416; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

QY 81 atggacaaaatagtaactcgttctgtctgtatctctgtctactcttctctagctctcaagg 140
   || || || || || || || || || || || || || || || || || || || || ||
Db 15 ACGAGCACCAATGTAAAGTCTGCCTTATGTCTCCAGCCNAITGCAAAAGGTCTCAAAAG 74

QY 141 tctatggccctgtgttcactctgtatcttggcctgaaaccccatagtgctgcatgcat 200
   || || || || || || || || || || || || || || || || || || || || ||
Db 75 TCTATGGTCTGTGTTCACCGTGTATTTGGCATGAATCCCATAGTGGTGTTCATGAT 134

QY 201 atgaagcagtgaaagagccctgattgattcttggagagagattttctgaaagagcattt 260
   || || || || || || || || || || || || || || || || || || || || ||
Db 135 ATGAGGAGGAGGAGGAGCCCTGATTGATAATGGAGAGGAGTTTCTGGAAGAGGCAAT 194

QY 261 tccactggctgaagagactaacagagatttggatgtgtagtgatgctgctgtcttccaca 320
   || || || || || || || || || || || || || || || || || || || || ||
Db 195 CCCCAATATCTCAAGAATTAATAAGGACTTGGTAGTGACATATTCTGTGTGAGCT 254

QY 321 tctgtcttgggagtgaggagatgaaacacagagacttacagagctctcctggcagagct 380
   || || || || || || || || || || || || || || || || || || || || ||
Db 255 TTGTTAACTGGGTGAGGGGATGGAACACAGAGCCCTAAAAGCTTCTC-AGCAGAGCT 313

QY 381 tggccatcacatagctgcccagtg--cagctctctcttcttgcctggagctcctcct 438
   || || || || || || || || || || || || || || || || || || || || ||
Db 314 TAGCTATCTGATGGCTGGCCGAGTGTTCAGCAGCTTCTTCCCTTGGCTGTGAAATCTCC 373

QY 439 cctagttctctctctctctgttaggaattgttttcagcaatgaaagaaatgaaagga 498
   || || || || || || || || || || || || || || || || || || || || ||
Db 374 CAGTTCTGCCCTTTTATAGGAATCATTTCCAGCAATGCAAGAGATGGAAGGA 433

QY 499 gatccggcgttctcctcctacgcctgcggaattttggagtgaggagagagcattga 558
   || || || || || || || || || || || || || || || || || || || || ||
Db 434 GATCCGGCGTTTCTCCCTCACAACTTCGGGAATTTGGGATGGGAAGAGAGGAGCATTGA 493

QY 559 gaaccgtgttcaagaggaagcccgctgcctgtgagagagttgagaaaaaaccaagg 614
   || || || || || || || || || || || || || || || || || || || || ||
Db 494 GGACCGTGTTCAGAGGAGGAGCTCACTGCTTGTGGAGGAGTTGAGAAAACCAAGG 549

RESULT 3
AV695514 631 bp mRNA EST 25-SEP-2000
LOCUS AV695514 GK Homo sapiens cDNA clone GKCBIB04 5', mRNA sequence.
DEFINITION AV695514
ACCESSION AV695514
VERSION AV695514.1 GI:10297377
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 631)
AUTHORS Wu, T., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.,
, Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang
, Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
```

```
JOURNAL      Unpublished (2000)
COMMENT      Contact: Zeguang Han
              Chinese National Human Genome Center at Shanghai
              351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
              201203, P. R. China
              Tel: 86-21-50801919(ex.45)
              Fax: 86-21-50801922
              Email: hanzg@chgc.sh.cn
              This clone is available at CHGC in Shanghai.

FEATURES     Location/Qualifiers
    source    1..631
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="GKCBID04"
              /tissue_type="hepatocellular carcinoma"
              /dev_stage="Adult"
              /lab_host="SOLR"
              /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
              XhoI"

BASE COUNT   164 a 128 c 160 g 178 t 1 others
ORIGIN

Query Match      43.4%; Score 322.8; DB 10; Length 631;
Best Local Similarity 77.6%; Pred. No. 4.7e-78;
Matches 416; Conservative 0; Mismatches 117; Indels 3; Gaps 2;

Oy 81 atgacacaaatagtaacttggttgctgttatctctgtctacttctcctagctctcaagg 140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 16 ACGAGCACCAATGTAAGTCGCTTATGTCCTCCAGCCCAATTGCCAAGGTTCTCAAAG 75

Oy 141 tctatgacctgttctcaactctgtatttgctgttatctctgtctacttctcctagctctcaagg 200
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 76 TCTATGGTCTGTGTTCCACCGTGATTTTGGCATGATTCCTCCATAGTGTTCTTCATGAT 135

Oy 201 atgaagcagtgaaagagccctgattgatcttgagagagagtttcttggaagagcattt 260
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 136 ATGAGGCAGTCAAGGAAGCCCTGATTGATAATGGAGAGGAGTTTCTTGGGAAGAGCAAT 195

Oy 261 tcccactggctgaagagctaaacagagatttgtagtgatgctgctgtcttcaagca 320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 CCCCAATATCTCAAGAAATTAATAAGGACTTTGGTAGTGACATATTTCTGTGTCAAGT 255

Oy 321 tctgtcttgaggatgggagatggaacacagagacttacagagctcctcgggagagcct 380
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 TTGGTAACCTGGGGTGAGGGGATGGAACACAGAGCCCTTAAAGCTTCTC-AGCAGAGCT 314

Oy 381 tggcccatccacatggctgcccagtg--cagcttctcttcttgcctggatctcct 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 TAGCCTATCTGCATGGCTGCCGAGTGTTGCAGACATTTCTTCTTGGCTGTGAATCTCC 374

Oy 439 cctagtttcgtttctctctctgttaggaattgtttcagcaatggaagaatggaagga 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 CAGTTTCTGCCCTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGGA 434

Oy 499 gatccggcgtttctcctctatgacgtcgcgaattttgggatgggagagagcattga 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 GATCCGGGTTTCTCCCTCACAACCTTGCAGAAATTTTGGGATGGGGAAGAGGACATIGA 494

Oy 559 ggaccgtgttcaagaggagcccgctgctgtgagagagttgagaaaaaccagg 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 GGACCGTGTTCAAGAGGAAGCTCACTGCCTTGTGGAGAGTTGAGAAAAACCAAG 550

RESULT 4
LOCUS      AV697749 592 bp mRNA EST 25-SEP-2000
DEFINITION AV697749 GK Homo sapiens cDNA clone GKCBIF10 5', mRNA sequence.
ACCESSION  AV697749
VERSION    AV697749.1 GI:10299612
KEYWORDS   EST.
SOURCE     human.
```

Qy:	499	gacccggcgtttctccctcatagacactcgcgaatttggagtgaggagagcattga	558
Db	420	GATCGGGGTTTCTCCCTCACAACCTTCGCGAATTTGGGATGNGAAGAGAGGACGATTTA	479
Qy:	559	gaaccgtttcaaaggaagccctcgtctgtgagagagttgagaaaaaaccaag	614
Db	480	GGACCGTGTTCAAGAGGAGGCTCACTGCTGTTGGAGGAGTTTGAAAAACCAAGG	535
RESULT	6		
LOCUS	AV696436		
DEFINITION	AV696436	675 bp	mRNA
ACCESSION	AV696436		EST
VERSION	AV696436.1	GI:10298299	25-SEP-2000
KEYWORDS	EST		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 675)		
	Wu, T., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H.		
	Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z., Zeng, L.		
	Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang		
	, Y., Gu, Y., Chen, Z. and Han, Z.		
TITLE	Homo sapiens CDNA GK- clones		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Zeguang Han		
	Chinese National Human Genome Center at Shanghai		
	351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai		
	201203, P. R. China		
	Tel: 86-21-50801919 (ex. 45)		
	Fax: 86-21-50801922		
	Email: hanzg@chgc.sh.cn		
FEATURES	This clone is available at CHGC in Shanghai.		
	Location/Qualifiers		
	1..675		
source	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		

```

/clone_rid="0"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT      178 a  136 c  162 g  198 t    1 others
ORIGIN

Query Matchbase 43.3%; Score 321.8; DB 10; Length 675;
Best Local Similarity 77.4%; Pred. No. 9e-78;
Matches 415; Conservative 0; Mismatches 118; Indels 3; Gaps 2

Qy      81 atggacaaaatagtaactcgtttctgtttattctgtctacttctctagctctcaaaag 140
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 ACGAGCACCAATGTAAAGTCGCTTATGTTCTCCAGGCCAATGCCAAGGTTCTCAAAAG 60

Qy     141 tctatggccctgtgttcaactcgtattttggcctgaaccccatagtgctgcgatggat 200
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     61 TCTATGGTCCTGTGTTCAACCGTGATTTTGGCATCAATCCCATAGTGGTGTTTCATGGAT 120

Qy     201 atgaagcagtgtaagaagacccctgatgatcttggaggagagtttctggaaagagcattt 260
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    121 ATGAGGCACGTGAAGGAAGCCCTGATTGATTAATGGAGAGGAGTTTTCTGGGAAGAGGCAATT 180

Qy     251 tccactggcgtgaagagcgtacacagagatttggtagtgatgctgcctgtttcacga 320
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db    181 CCCCATATCTCAAGAATTACTAAGGACCTTGGTAGGTGCACATATTCTGTGTGACGCT 240

Qy     321 tctgtcttgggagtgaggagtgaaacacagacattacagagctcctcgggcagagct 380
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db	241	TTGGTAAC	TGGGGT	GAGGGG	ATGGAAA	CACAGACG	CTTAA	AAAGCTTCTC	-AGCAGAGCT	299
QY	381	tgcccatccac	atgctgctgcccagtg-	-cagcttctctcttcttctgctggaatctccct	438					
Db	300	TAGCCTATCTG	CATGGCTGCGAGTG	TGCAGCAGCACTTCTCTCTGGCTGTCAA	11CTCC	359				
QY	439	cttagttcgtttctctctctgtaggaattgtttcagcaaatgaaagaatgaaaga	498							
Db	360	CAGTTTCTG	CGCCCTTTTTTTTATAGGAATCA	TTCCAGCAATGGAAGAGATGA	GAAGGA	419				
QY	499	gacccggcgtttctctctctcctcctcctgacgcgtgcggaattttggatgggaaagagcattga	558							
Db	420	GATCCGGCGTTCTCCCTCACACCTTGC	CGGAATTTGGATGNGAAGAGGAGCA	TTCGA	479					
QY	559	ggaccgtgttcaagagaagcccgctgccttctgtaggaggttgagaaaaaccaa	614							
Db	480	GGACCGTGTTC	CAAGAGGAGCTACCTGCCTTGTGGAGGAGTTGAGAAAAACCAAG	535						
RESULT	7									
AV697538										
LOCUS		AV697538	GKC Homo sapiens	cdna clone KCBJB03	5', mRNA sequence.	EST	25-SEP-2000			
DEFINITION		AV697538								
ACCESSION		AV697538.1	GI:10299401							
VERSION		EST.								
KEYWORDS		human.								
SOURCE		Homo sapiens								
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
REFERENCE		1 (bases 1 to 677)								
AUTHORS		Wu, T., Qian, B., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L., Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.								
TITLE		Homo sapiens cDNA GK- clones								
JOURNAL		Unpublished (2000)								
COMMENT		Contact: Zeguang Han Chinese National Human Genome Center at Shanghai 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China Tel: 86-21-50801919(ex.45) Fax: 86-21-50801922 Email: hanzg@chgc.sh.cn This clone is available at CHGC in Shanghai.								
FEATURES		Location/Qualifiers								
source		1..677								
		/organism="Homo sapiens"								
		/db_xref="taxon:9606"								
		/clone="KCBJB03"								
		/clone_lib="GKC"								
		/tissue_type="hepatocellular carcinoma"								
		/dev_stage="Adult"								
		/lab_host="SOLR"								
		/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"								
BASE COUNT		178 a 141 c 165 g 191 t								
ORIGIN		2 others								
Query Match		43.3%								
Best Local Similarity		77.4%								
Matches	415;	Conservative	0;	Mismatches	118;	Indels	3;	Gaps	2	
QY	81	atggacaaaata	gaactcgtttgctgtattctctgtctacttctcctagctctcaaga	140						
Db	2	ACGAGCACC	AATGTAAGTCGCTTATGTCTCTCAGCAATTGCAAGGTTCTCAAAG	61						
QY	141	tctatggccctggttcactctgtattttggcctgaaccataagtgctgcctagat	200							
Db	62	TCTAAGTCTGTGTTT	CAACGGTGTATTTTGGCAIGAAATCCCATAGTGTGTTCATGAT	121						
QY	201	atgaagcagtgaaagaaacccctgattgactcttggaggagagtttcttggaaagagcatt	250							

[illegible]

Query Match	43.3%	Score 321.8;	DB 10;	Length 677;
Best Local Similarity	77.4%;	Pred. No. 9e-78;		
Matches 415;	Conservative	0;	Mismatches 118;	Indels 3;
Gaps 2				

```

QY 81 atggacaaaatagtaactcgttctgtgtatctctgtctatctctctctctctctctctctcaagg 140
Db 2 ACAGCACCACCAATGTAAGTCTGCTTATGTTCTCCAGGCAATTCGAAAGGTTCTCAAAAG 61

QY 141 tctatggccctgttctcactctgtatttttggcctgaaaccccatagtgctgctgcacgat 200
Db 62 TCTATGGTCTGTGTTCCACCGTGATTTTGGCATGAATCCCATAGGTGTTTCAAGAT 121

QY 201 atgaacagtgaaaggaagccctgattgatcttggagagagatttcttggaaagagcattt 260
Db 122 ATGAGGCAAGTGAAGGAGCCCTCAITGATAATGGAGAGAGTTTCTGGAAGAGGCAAT 181

QY 261 tccactggctaaaagactaaacagagagatttggtagtgatgcattgcctcttttcagca 320
Db 182 CCCAATACTCAAGAAATTAATAAGGACTTGGTAGGTGCACATATTTCTGTGTCAGCT 241

QY 321 tctgtcttgggagtgaggagagtggaacacagagacttacagagctcctcgggagagct 380
Db 242 TTGTAACCTGGGGTGAGGGGATGGAACACAGAGCCCTAAAGAGCTTCTC-AGCAGAGCT 300

QY 381 tggccatcacatagctgcccagtg--cagcttctcttcttcttcttcttcttcttcttct 438
Db 301 TAGCCTATCTGCAAGCTGCCAGTGTGCAGCATTCTTCTTCTGCTGTGAAATCTCTC 360

QY 439 cctagttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 498
Db 361 CAGTTCTGCCCCCTTTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGA 420

QY 499 gatccggcgtttctcctcatgacgctgcggaatttggatgggaaagagagacattga 558
Db 421 GATCCGGCGGTTCTCCCTCACAACTTGGCGAATTTGGGATGGNCAAGAGAGCATTCA 480

QY 559 ggaccgtgttcaagaggaagcccgctgtcttggagagagtgagaaaaacccaagg 614
Db 481 GGACCGGTGTTCAAGAGGAAGCTCACTGCCTTGTGGAGGAGTTGAGAAAAACCAAGS 536

RESULT 9
AV696639 580 bp mRNA EST 25-SEP-2000
LOCUS AV696639 GK Homo sapiens cDNA clone GKCBIE09 5', mRNA sequence.
DEFINITION AV696639
ACCESSION AV696639
VERSION AV696639.1 GI:10298502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHOR Wu, T., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang,
Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Location/Qualifiers
1..680
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCBIE09"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

```

```

XhoI"
BASE COUNT 177 a 141 c 167 g 192 t 3 others
ORIGIN

Query Match 42.7% Score 317: DB 10: Length 680:
Best Local Similarity 76.9% Pred. No. 1.9e-76:
Matches 412: Conservative 0: Mismatches 121: Indels 3: Gaps 2:

QY 81 atggacaaaatagtaactcgttctgtgtatctctgtctatctctctctctctctctcaagg 140
Db 11 ACGAGCACCACCAATGTAAGTCTGCTTATGTTCTCCAGGCAATTCGAAAGGTTCTCAAAAG 70

QY 141 tctatggccctgttctcactctgtatttttggcctgaaaccccatagtgctgctgcacgat 200
Db 71 TCTATGGTCTGTGTTCCACCGTGATTTTGGCATGAATCCCATAGGTGTTTCAAGAT 130

QY 201 atgaacagtgaaaggaagccctgattgatcttggagagagatttcttggaaagagcattt 260
Db 131 ATGAGGCAAGTGAAGGAGCCCTGATTGATAATGGAGAGAGTTTCTGGAAGAGGCAAT 190

QY 261 tccactggctgaaagagactaacagagagatttggtagtgatgcattgcctcttttcagca 320
Db 191 CCCAATACTCAAGAAATTAATAAGGACTTGGTAGGTGCACATATTTCTGTGTCAGCT 250

QY 321 tctgtcttgggagtgaggagagtggaacacagagacttacagagctcctcgggagagct 380
Db 251 TTGTAACCTGGGGTGAGGGGATGGAACACAGAGCCCTAAAGAGCTTCTC-AGCAGAGCT 309

QY 381 tggccatcacatagctgcccagtg--cagcttctcttcttcttcttcttcttcttcttct 438
Db 310 TAGCCTATCTGCAAGCTGCCAGTGTGCAGCATTCTTCTTCTGCTGTGAAATCTCTC 369

QY 439 cctagttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttctt 498
Db 370 CAGTTCTGCCCCCTTTTTTATTAGGAATCATTTCCAGCAATGGAAGAGATGGAAGA 429

QY 499 gatccggcgtttctcctcatgacgctgcggaatttggatgggaaagagagacattga 558
Db 430 GATCCGGCGGTTCTCCCTCACAACTTGGCGAATTTGGGATGGNCAAGAGAGCATTGA 489

QY 559 ggaccgtgttcaagaggaagcccgctgtcttggagagagtgagaaaaacccaagg 614
Db 490 GGACCGGTGTTCAAGAGGAAGCTCACTGCCTTGTGGAGGAGTTGAGAAAAACCTCAAG 545

RESULT 10
AV697579 622 bp mRNA EST 25-SEP-2000
LOCUS AV697579 GK Homo sapiens cDNA clone GKCBJB12 5', mRNA sequence.
DEFINITION AV697579
ACCESSION AV697579
VERSION AV697579.1 GI:10299442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHOR Wu, T., Qian, B., Huang, Q., Kang, B., Gao, X., Xu, Z., Xiao, H.,
Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L.,
Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang,
Y., Gu, Y., Chen, Z., and Han, Z.
Homo sapiens cDNA GK- clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
Location/Qualifiers

```

Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

Location/Qualifiers
1..747
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCATD11"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 212 a 141 c 186 g 206 t 2 others
ORIGIN

Query Match 39.0%; Score 289.6; DB 10; Length 747;
Best Local Similarity 98.6%; Pred. No. 5.5e-69;
Matches 292; Conservative 0; Mismatches 4; Indels 0; Gaps

Oy 1 tcagaaatattgaagcctgttgctgataaaaagcatacaaatgaaatatca 60
Db 192 TCAGAAATATTGAAGCCTGTGGTGAATAAAGCATACAAATGAAAATATCA 251
Oy 61 tgctaactcaggcttagcaaatgacaaaatagttaacctgtttgcttatctgtct 120
Db 252 TGTAAATCAGCGTTACCAATAGAACAAATAGTAACCTTCGTTTGCTATCTCTGCT 311
Oy 121 acttctagctctcaaggctctatggccctgtgttcactctgtatttggcctgaacc 180
Db 312 ACTTCTAGCTCTCAAAGGCTATGGCCCTGTGTTCACTCTCTATTGGCCTGAAACC 371
Oy 181 catagtggtcatggtatgaacagcagtgaaagcctgatgatctgagaga 240
Db 372 CATAGTGGTCTGATGATGAAGCAGTAGAAGGAGGCCCTGATTGATCTTGAGAGGA 431
Oy 241 gtttctggaagaggcattttcccactggctgaaagagctaacagaggatttgta 296
Db 432 GTTTCCTGGAAGGCGATTTCCTCACTGGCTGAAAGAGCTAACAGAGGATTGGA 487

RESULT 12
AV656485
LOCUS AV656485 646 bp mRNA EST 25-SEP-2000
DEFINITION AV656485 GRK Homo sapiens CDNA clone KCBJAL2 5', mRNA sequence.
ACCESSION AV656485
VERSION AV656485.1 GI:10298348
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 646)
AUTHORS Wu,T., Qian,B., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,
G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
YU, Gu.Y., Chen,Z. and Han,Z.
Homo sapiens CDNA GK-clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn

1..622
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GKCJB12"
/clone_lib="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 165 a 126 c 156 g 175 t
ORIGIN

Query Match 42.4%; Score 314.8; DB 10; Length 622;
Best Local Similarity 76.7%; Pred. No. 7.5e-76;
Matches 411; Conservative 0; Mismatches 122; Indels 3; Gaps 2;

Oy 81 atgacaaaatagtaactctgttgcgtgtatctctgtctacttctcctagctccaagg 140
Db 1 ACGAGCCAATGATGATGCTTATGTTCTCACGCCAATGCCAAGGTTCTCAAAAG 50
Oy 141 ttctatggccctgtttcacctctgtattttggcctgaaacccatagtggtgctcgatgat 200
Db 61 TCTATGTCCTGTGTTCACCGCTGATTTTGCATGATCCCATAGTGGTGTTCATGAT 120
Oy 201 atgaagcagtaagaagcctgattgattctgtggagagagatttctggaagagcattt 260
Db 121 ATGAGGCAGTGAACGAAGCCCCTGATTGATAATGAGAGGAGGTTCTTGGAGAGGCCAAT 180
Oy 261 tcccactggctgaaagagctaacagagagatttgtagtgtgcagtgttcctgttcagca 320
Db 181 CCCCAATATCTCAAGAATACTTAAGGACTTGTAGTGCACATATTTCGTGTCAGCT 240
Oy 321 tctgtctgggagtggggagatgaaacacagagacttacagagctcctcgggcagagct 380
Db 241 TTGTTAAGTGAAGTGAGGGCGATGGAAAAAGAGCCCTAAAAGCTTCTC-AGCAGAGCT 299
Oy 381 tggccatccacatggctgccaggt--cagcttctcttcttctgctgggactctccct 438
Db 300 TAGCCTATCTGCATGCTGCCAGTGTTCGAGCACTTCTTCCTTGGCTGTGATTTCC 359
Oy 439 cctagttcgttctctctctgttagaattgttttcagcaatggaagaatggaagga 498
Db 360 CAGTTTCGCCCTTTTTTATAGGAATCAITTCAGCAATGACAGAGATGGAAGGA 419
Oy 499 gatccggcgttctccctcatgacgctgcggaatttggagtgagggaagagagcattga 558
Db 420 GATCGGGGTTTCTCCCTCACACCTTCGGGAATTTGGGATGGGAGGAGGAGCATIGA 479
Oy 559 ggaccgtgttcaagaagaagcccgctctgtgtggagaggttgaaaaaccaagg 614
Db 480 GGACCGTGTTCAGAGGAAGCTCACTGCCCTGTGGAGGAGTTGAGAAAAACCAAG 535

RESULT 11
AV646984
LOCUS AV646984 747 bp mRNA EST 07-SEP-2000
DEFINITION AV646984 GLC Homo sapiens CDNA clone GLCATD11 3', mRNA sequence.
ACCESSION AV646984
VERSION AV646984.1 GI:9867998
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 747)
AUTHORS Qian,B., Wu,T., Huang,Q., Huang,C., Kang,B., Gao,X., Xu,Z., Xiao,H.,
Qian,B., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Qu,J.,
Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu
G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
Homo sapiens CDNA clone
Unpublished (2000)

This clone is available at CHGC in Shanghai.

FEATURES

source

Contact: Robert Strausberg, Ph.D.
Email: cchapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LILCM1628 row: b column: 11
High quality sequence stop: 718.
Location/Qualifiers
1..721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766890"
/clone_lib="NIH_MGC_76"
/lab_host="DH10B (T1 phage-resistant)"
/note="Jordan: Liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccattcgcc); Site_2: SfiI (ggccattcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGAGCGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

176 a 147 c 195 g 203 t

Query Match 36.4%; Score 270.4; DB 11; Length 721;
Best Local Similarity 99.3%; Pred. No. 1.2e-63;
Matches 282; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 461 ttaggcaattgtttcagcaatggaagaatggaaggagatccggcggtttctccctcag 520
Db 355 TTTCGAATTTGTTTTCAGCAATGGAAGAAATGGAAGAGATCCGGCGTTTCTCCCTCATG 414
Qy 521 acgtctcggaatttggatggggaaggagcattgagaccgtgttcaagagaagcc 580
Db 415 ACCTGCGGAAATTTTGGGATGGGGAAGAGACATTTGAGACCGTGTTCAGAGAGAGACC 474
Qy 581 cgtctccttggaggaggttgagaaacaaaggggtgggtgacctactccatcactg 640
Db 475 CGCTGCTTGTGAGGAGTGTGAGAAACCAAGGTTGGTGTGACCTTACTTATCATCTG 534
Qy 641 accttactggactactcttcttactgacattcttgaacatttcaggggtggccat 700
Db 535 ACCTTACTGGACTACTATCTTCTTACTGACATCTTGTGAACATTTTCAGGGGTGGCCAT 594
Qy 701 atcttctattatgagt-ctgggtttagctcatgtgaagcgggg 743
Db 595 ATCTTTTCAITATGAGTCCTGGTGTGTAGTCTATGTGTAAGCGGGG 638

RESULT 14

AV595838 530 bp mRNA EST 25-SEP-2000
LOCUS
DEFINITION AV595838 GKc Homo sapiens cDNA clone GKCBEE06 5', mRNA sequence.
ACCESSION AV595838
VERSION AV595838.1 GI:10297701
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 530)
Wu, J., Qian, B., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z., Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Ou, J., Song, H., Cheng, Z., Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.

BASE COUNT 172 a 130 c 161 g 183 t

ORIGIN

Query Match 37.7%; Score 280.2; DB 10; Length 646;
Best Local Similarity 72.6%; Pred. No. 2.5e-66;
Matches 390; Conservative 0; Mismatches 143; Indels 4; Gaps 2;
Qy 81 atggacaaaatagtaacttcgttggctgttatctgtctacttctcagctctcaagg 140
Db 2? ACGAGCAACCAATTAAGTCTGCCTTATGTTCTTACAGCCAAATTCGAAAGGTTCTCAAAAG 81
Qy 141 tctatggccctgttctcactctgtatttggcctgaaacccatagtggtctcagcagat 200
Db 82 TCTATGCTCTGTGTTACCGCTGATTTTGGCATGAATTCATAGAGGTGTTTATGGAT 141
Qy 201 atgaagcagtgaaaggacccctgattgatcttggaggagagatttctgaaaggcattt 260
Db 142 ATGAGGCAGTGAACGAAGCCCTGATTGATGATGAGAGTAGGTCCTTGGGAAGGCAATTT 201
Qy 261 tccactggctgaaagcttaacagagatttggtagtgatgctgctcttcttcagca 320
Db 202 ACCCAATATCTCAAGAAATTAATTAAGGCTTGGTAGGTGCACATATATCTGTGTCAGCT 261
Qy 321 tctgtcttggggatgggagatggaacacagagacttacagagctcctcggagcagct 380
Db 262 TTGTAACCTGGAGTGGGGCGATGGAACACAGAGCCCTAAAGAGCTAGTCT-ATCAGATCT 320
Qy 381 tggccatcacatggctgccagtgctcagcttctctt-cttgcctggagctccc 437
Db 321 TATCTATCTGCAATGATGCGAGTGTGTGATCATTCTTACCCTGGCTGTGNAATATCTC 380
Qy 438 tctagttctgttcttctccttctaggaattgttttcagcaatggaagaaatggaagg 497
Db 381 CCAGATTCTGCCCTGAAATTTTATTAGGAATGATTTCCAGCAATGGACAGACATGGAAG 440
Qy 498 agatccggcgttctcctcctacgcctgcggaatttgggatggggaagagacatig 557
Db 441 AGATCCGGGCTTCTCCTCACACCTTGGCGGAATTTGGGATGGGAGAGAGCATIG 500
Qy 558 aggaccgtgttcaagaggagcccgctgcttgtgagagaggttgagaaaccccaagg 514
Db 501 AGGACCGTTTTCAAGAGAGAGCTCACTGCCTTGTGCGGAGTTGCGAAACCAAGG 557

RESULT 13

BG617892 721 bp mRNA EST 18-APR-2001
LOCUS
DEFINITION BG617892 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4765890 5', mRNA sequence.
ACCESSION BG617892
VERSION BG617892.1 GI:13669263
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
REFERENCE
AUTHORS
1 (bases 1 to 721)
National Institutes of Health, Mammalian Gene Collection: (MGC)
JOURNAL Unpublished (1999)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 19, 2002, 08:10:30 : Search time 130.44 Seconds
(without alignments)
1290.043 Million cell updates/sec

Title: US-09-763-292-2
Perfect score: 743
Sequence: 1 tcagaaatattgaagcctg.....ttagctcatgtgaacgagg 743

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS-COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	163	21.9	1419	2	US-08-194-981E-4
2	163	21.9	1591	2	US-08-194-981E-3
3	163	21.9	1852	1	US-08-201-118-10
4	163	21.9	1852	2	US-08-238-821B-10
5	163	21.9	1852	5	PCT-US95-05744-10
6	163	21.9	1854	1	US-08-201-118-4
7	163	21.9	1854	2	US-08-238-821B-4
8	163	21.9	1854	5	PCT-US95-05744-4
9	153.4	20.6	1892	1	US-08-201-118-14
10	153.4	20.6	1892	2	US-08-238-821B-14
11	153.4	20.6	1892	5	PCT-US95-05744-14
12	149.6	20.1	1746	1	US-08-201-118-2
13	149.6	20.1	1746	2	US-08-238-821B-2
14	149.6	20.1	1746	5	PCT-US95-05744-2
15	144	19.4	200	1	US-08-716-459-13
16	140.2	18.9	2009	1	US-08-201-118-6
17	140.2	18.9	2009	2	US-08-238-821B-6
18	140.2	18.9	2009	5	PCT-US95-05744-6
19	140.2	18.9	2258	1	US-08-201-118-12
20	140.2	18.9	2258	2	US-08-238-821B-12
21	140.2	18.9	2258	5	PCT-US95-05744-12
22	134.2	18.1	1829	1	US-08-201-118-8
23	134.2	18.1	1829	2	US-08-238-821B-8
24	134.2	18.1	1829	5	PCT-US95-05744-8
25	89.4	12.0	100	1	US-08-716-459-12
26	79.2	10.7	1419	2	US-08-194-981E-2
27	79.2	10.7	1482	2	US-08-194-981E-1

Query Match 21.9% Score 163; DB 2: Length 1419;
Best Local Similarity 97.1%; Pred. No. 1.4e-42;

ALIGNMENTS

RESULT 1

US-08-194-981E-4

: Sequence 4, Application US/08194981E

: Patent No. 5886157

: GENERAL INFORMATION:

: APPLICANT: GUENGERICH, F. Peter

: APPLICANT: GUO, Zuyu

: APPLICANT: SANDHU, Punam

: APPLICANT: GILLIAM, Elizabeth M. J.

: TITLE OF INVENTION: EXPRESSION AND PURIFICATION OF

: TITLE OF INVENTION: HUMAN

: TITLE OF INVENTION: CYTOCHROME P450

: NUMBER OF SEQUENCES: 68

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: NEEDLE & ROSENBERG, P.C.

: STREET: Suite 1200, 127 Peachtree Street, NE

: CITY: Atlanta

: STATE: Georgia

: COUNTRY: USA

: ZIP: 30303-1811

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: PatentIn Release #1.0, Version #1.25

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/194,981E

: FILING DATE: February 10, 1994

: CLASSIFICATION: 435

: ATTORNEY/AGENT INFORMATION:

: NAME: Elizabeth Selby

: REGISTRATION NUMBER: 38,298

: REFERENCE/DOCKET NUMBER: 22000.0022

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (404) 688-0770

: TELEFAX: (404) 688-9880

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 1419 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: DNA (genomic)

: HYPOTHETICAL: NO

: ANTI-SENSE: NO

: US-08-194-981E-4

Sequence 3, Appli
Sequence 4, Appli
Sequence 2, Appli
Sequence 1, Appli
Sequence 11, Appli
Sequence 20, Appli
Patent No. 5508199
Sequence 13, Appli
Sequence 22, Appli
Sequence 23, Appli
Sequence 21, Appli
Sequence 24, Appli
Sequence 1, Appli
Sequence 14, Appli
Sequence 45, Appli
Sequence 45, Appli
Sequence 14, Appli

Sequence 10, Application US/08238821B
Patent No. 5912120
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
APPLICANT: DE MORAIS, Sonia M.F.
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT OF S-
TITLE OF INVENTION: MEPHENYTOIN METABOLISM
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-192110US
TELEPHONE: (650) 326-2400
TELEFAX: (650) 326-2422
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Region
LOCATION: 1..10
OTHER INFORMATION: /note= "Corresponds to positions -10 to -1
OTHER INFORMATION: for 55 of Figure 2."

US-08-238-821B-10

Query Match 21.9%; Score 163; DB 2; Length 1852;
Best Local Similarity 97.1%; Pred. No. 1.6e-42;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 126 cctagctctcaaggtctatggccctgtgttcactctgtattttggcctgaaacccatag 185
|||
DB 174 CCAATCTCTCAAAAGGCTATGAGCAGTGAAGGAGCCCTGATCTGTATTTGGCCTGAAACCCATAG 233
QY 186 tggctctcgtatgataaagcagtgaaagcctgattgattcttgaggagatttt 245
|||
DB 234 TGGTGTCTGATGATGAAGCAGTGAAGGAGCCCTGATCTGTATTTGGCCTGAAACCCATAG 293
QY 246 ctggaagagcattttccctcactggctgaaagagcctaaacagagatttgta 296
|||
DB 294 CTGGAAGAGGCAATTTTCCCACTGGCTGAAAGAGCTAAACAGAGGATTITGGAA 344

RESULT 5

PCT-US95-05744-10
Sequence 10, Application PC/TUS9505744
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
APPLICANT: DE MORAIS, Sonia M.F.
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
TITLE OF INVENTION: OF S-MEPHYNYTOIN METABOLISM
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05744
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,821
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,962
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 15280-192-1-1
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1852 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-05744-10

Query Match 21.9%; Score 163; DB 5; Length 1852;
Best Local Similarity 97.1%; Pred. No. 1.6e-42;
Matches 166; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 126 cctagctctcaaggtctatggccctgtgttcactctgtattttggcctgaaacccatag 185
|||
DB 174 CCAATCTCTCAAAAGGCTATGAGCAGTGAAGGAGCCCTGATCTGTATTTGGCCTGAAACCCATAG 233
QY 186 tggctctcgtatgataaagcagtgaaagcctgattgattcttgaggagatttt 245
|||
DB 234 TGGTGTCTGATGATGAAGCAGTGAAGGAGCCCTGATCTGTATTTGGCCTGAAACCCATAG 293
QY 246 ctggaagagcattttccctcactggctgaaagagcctaaacagagatttgta 296
|||
DB 294 CTGGAAGAGGCAATTTTCCCACTGGCTGAAAGAGCTAAACAGAGGATTITGGAA 344

RESULT 6

US-08-201-118-4
Sequence 4, Application US/08201118
Patent No. 5786191

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/864,952
FILING DATE: 09-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Liebeschuetz, Joe
REGISTRATION NUMBER: 37,505
REFERENCE/DOCKET NUMBER: 15280-19211005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 326-2400
TELEFAX: (650) 326-2422
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1854 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: Region
LOCATION: 1..12
OTHER INFORMATION: /note= "Corresponds to position 25 of Figure 2."
US-08-238-821B-4

[illegible]

RESULT 8
 PCT-US95-05744-4
 : Sequence 4, Application PC/TUS9505744
 : GENERAL INFORMATION:
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : APPLICANT: DE MORAIS, Sonia M.F.
 : TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
 : TITLE OF INVENTION: CYTOCHROME P450 2C19: THE PRINCIPAL DETERMINANT
 : TITLE OF INVENTION: OF S-MEPHNYTOIN METABOLISM
 : NUMBER OF SEQUENCES: 61
 : CORRESPONDENCE ADDRESS:

```

Query Match      20.6%; Score 153.4; DB 1; Length 1892;
Best Local Similarity 91.8%; Pred. No. 2e-39;
Matches 157; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
Q: 126 cctagctctcaaaaggtctatggccctgtgttcactctgtatttggcctgaaccccatag 185
    |||

```

	Query Match	20.6%	Score 153.4	DB 1	Length 1892
	Best Local Similarity	91.8%	Pred. No. 2e-39		
	Matches 157	Conservative	0	Mismatches 14	Indels 0
Q:	126	cctagctctcaagaggtctatggccctgtgttcactctgtatttggccctgaaccccatag	185		
Db	206	CCAATNTCTCAAAAGTCTATGGCCCTGTGTTCACNTGTATTTTGGCTGAAGCCCATAG	265		
Q:	185	tgttgctcatgatatagaagcagatgaagaagccctgattgatctctggaggaagatttt	245		
Db	266	TGTTGNTGCAATGATATGAAGACGTGAAGGAAGCCCTGATTGATCNTGGAGAGAGATTTT	325		

Q: 246 ctggaagaggcattttccactgctgaaagctaacagagatttgta 296
|||||
D: 326 CTGGAAGAGGCANTTTCCTGCTGAAAGAGNTAACANAGGANTTGA 376
|||||

APPLICANT: DE MORAIS, SOLIA M.F.
TITLE OF INVENTION: CLONING. EXPRESSION AND DIAGNOSIS OF HUMAN

RESULT 10
US-08-238-821B-14
; Sequence 14, Application US/08238821B
; Patent No. 5912120

CORRESPONDENCE ADDRESS:

: ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

: CITY: San Francisco

STATE: California

; COUNTRY: US

ZIP: 94111

: : COMPUTER READABLE FORM:

```

:
: MEDIUM TYPE: Floppy disk
:
COMPUTER  IBM PC

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: DOS 4.0 or DOS 5.0

OPERATING SYSTEM: PC-DOS/MS-DOS

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCI/US95/05744
: FILING DATE:
: CLASSIFICATION:
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/238,821
: FILING DATE: 06-MAY-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/201,118
: FILING DATE: 22-FEB-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/864,962
: FILING DATE: 09-APR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Dow, Karen B.
: REGISTRATION NUMBER: 29,684
: REFERENCE/DOCKET NUMBER: 15280-192-1-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 326-2400
: TELEFAX: (415) 326-2422
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1892 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: YES
:
: PCI-US95-05744-14

```

	Query Match	20.6%	Score 153.4	DB 5	Length 1892
	Best Local Similarity	91.8%	Pred. No. 2e-39		
	Matches 157	Conservative 0	Mismatches 14	Indels 0	Gaps 0
Qy	126	cctagctctcaaaaggtctatagccctggttcactgtatctgtatttggccctgaaccccatag	185		
Db	206	CCAATNTCTCAAAAGTCTATGGCCCTGTGTCACNTGTATTTTGGCCCTGAACCCCATAG	265		
Qy	186	tggTgctcattggtatgaagcaagtgaaggagccctgattgactcttgtaggagagatttt	245		
Db	256	TGGTNTGCATTGGATATGAACAGTGAAGGAAGCCCTGATTGATCNTGGAGGAGT	325		
Qy	246	ctggaagagggcattttccctcactggctgaaagagctaaacagagatttgta	296		
Db	376	CTGAAGAGGCANTTTCCTCTGCTGAAGAGNTAACNAGGANTTGGAA	376		

```

RESULT 12
US-08-201-118-2
: Sequence 2, Application US/08201118
: Patent No. 5786191
: GENERAL INFORMATION:
: APPLICANT: GOLDSTEIN, Joyce A.
: APPLICANT: ROMKES-SPARKS, Marjorie
: TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLE
: TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE
: TITLE OF INVENTION: SUBFAMILY
: NUMBER OF SEQUENCES: 44
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Townsend and Townsend Khourie and Crew
: STREET: 379 Lytton Avenue
: CITY: Palo Alto
: STATE: California
: COUNTRY: US
: ZIP: 94301
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25

```

RESULT 11
 PCT-US95-05744-14
 : Sequence 14, Application PC/TUS9505744
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: GOLDSTEIN, Joyce A.
 : APPLICANT: ROMKES-SPARKS, Marjorie
 : TITLE OF INVENTION: CLONING AND EXPRESSION OF COMPLEMENTARY
 : TITLE OF INVENTION: DNAS FOR MULTIPLE MEMBERS OF THE HUMAN CYTOCHROME P450 2C

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/201,118
;; FILING DATE: 22-FEB-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/864,962
;; FILING DATE: 09-APR-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Liebeschuetz, Joe
;; REGISTRATION NUMBER: 37,505
;; REFERENCE/DOCKET NUMBER: 15280-192-1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 326-2400
;; TELEFAX: (415) 326-2422
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1746 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-08-201-118-2

Query Match 20.1%; Score 149.6; DB 1; Length 1745;
Best Local Similarity 97.4%; Pred. No. 3.2e-38;
Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 461 ttaggaattgttttcagcaatgaaagaaatgaaagagatccgcggtttctccctcatg 520
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 333 TTGGAAATCGTTTTCAGCAATGAAAGAGATGGAAGAGATCCGGCGTTTCTCCCTCATG 392

QY 521 acgctcggaattttggatggggaagagagagcattgagaccgttttcaagaggaagcc 580
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 393 ACCTCGCGAAATTTGGGATGGGAAGAGAGCATTGAGACCGTTTCAAGAGGAAGCC 452

QY 581 cgcctcctgtgagaggttgagaaacccaaggtt 616
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 453 CGCTGCCITTTGGAGGAGTTGAGAAAAACCAAGGCT 488

RESULT 13
US-08-238-821B-2
Sequence 2, Application US/08238821B
Patent No. 5912120
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
APPLICANT: DE MORAIS, Sonia M.F.
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19; THE PRINCIPAL DETERMINANT OF S-
TITLE OF INVENTION: MEPHENYTOIN METABOLISM
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,821B
FILING DATE: 06-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/201,118
FILING DATE: 22-FEB-1994
CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/864,962
;; FILING DATE: 09-APR-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Liebeschuetz, Joe
;; REGISTRATION NUMBER: 37,505
;; REFERENCE/DOCKET NUMBER: 15280-192110US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (650) 326-2400
;; TELEFAX: (650) 326-2422
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1746 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: Region
;; LOCATION: 1..5
;; OTHER INFORMATION: /note= "Corresponds to positions -5 to-1
;; OTHER INFORMATION: for 11a of Figure 2."
US-08-238-821B-2

Query Match 20.1%; Score 149.6; DB 2; Length 1746;
Best Local Similarity 97.4%; Pred. No. 3.2e-38;
Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 461 ttaggaattgttttcagcaatgaaagaaatgaaagagatccgcggtttctccctcatg 520
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 333 TTGGAAATCGTTTTCAGCAATGAAAGAGATGGAAGAGATCCGGCGTTTCTCCCTCATG 392

QY 521 acgctcggaattttggatggggaagagagcattgagaccgttttcaagaggaagcc 580
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 393 ACCTCGCGAAATTTGGGATGGGAAGAGAGCATTGAGACCGTTTCAAGAGGAAGCC 452

QY 581 cgcctcctgtgagaggttgagaaacccaaggtt 616
|||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 453 CGCTGCCITTTGGAGGAGTTGAGAAAAACCAAGGCT 488

RESULT 14
PCT-US95-05744-2
Sequence 2, Application PC/TUS9505744
GENERAL INFORMATION:
APPLICANT: GOLDSTEIN, Joyce A.
APPLICANT: ROMKES-SPARKS, Marjorie
APPLICANT: DE MORAIS, Sonia M.F.
TITLE OF INVENTION: CLONING, EXPRESSION AND DIAGNOSIS OF HUMAN
TITLE OF INVENTION: CYTOCHROME P450 2C19; THE PRINCIPAL DETERMINANT
TITLE OF INVENTION: OF S-MEPHENYTOIN METABOLISM
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Khourie and Crew
STREET: 379 Lytton Avenue
CITY: Palo Alto
STATE: California
COUNTRY: US
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/05744
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/238,821
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:

